AD_____

Award Number: DAMD17-00-1-0050

TITLE: The Prostate Expression Database (PEDB)

PRINCIPAL INVESTIGATOR: Peter S. Nelson, M.D.

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, WA 98109-1024

REPORT DATE: April 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2003	Final (1 Apr 2000 - 31 Mar 2003)

4. TITLE AND SUBTITLE
The Prostate Expression Database (PEDB)

5. FUNDING NUMBERS
DAMD17-00-1-0050

6. AUTHOR(S)
Peter S. Nelson, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fred Hutchinson Cancer Research Center
Seattle, WA 98109-1024

E-Mail: pnelson@fhcrc.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: ALL DTIC reproductions will be in black and white

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This proposal aims to exploit advances in biotechnology and informatics to develop a genetics resource termed the Prostate Expression Database (PEDB) (http://www.pedb.org). The foundation of PEDB is the identification and characterization of a prostate transcriptome, the intermediary between the genome and the proteome that represents that portion of the human genome actively used or transcribed in the prostate. The research accomplished to date has assembled a working virtual prostate transcriptome that defines the genes used or transcribed in prostate cell types and tissues. This transcriptome has a physical counterpart of >10,000 cDNAs arrayed in cDNA microarray format for large-scale expression studies. This transcriptome has been used as a foundation for studies of the prostate proteome, the working counterpart to the genome and transcriptome. Our results show that these approaches are complementary. Analysis of the virtual transcriptome of LNCaP cells has identified 15 new androgen-regulated genes to date. We have extended PEDB to include sequence analysis of the murine prostate and constructed a corresponding database, mPEDB to facilitate the dissemination of mouse prostate gene expression information. The resources generated by the accomplishments of the objectives have utility in identifying genetic alterations in human and mouse prostate carcinoma that can be exploited as diagnostic or therapeutic targets.

14. SUBJECT TERMS
Prostate Cancer, transcriptome, cDNA, database

15. NUMBER OF PAGES
82

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

The objectives of this research proposal were to exploit advances in biotechnology and informatics to develop a genetics resource termed the Prostate Expression Database (PEDB) (http://www.pedb.org). The PEDB was envisioned as an integrated resource focused exclusively on prostate cancer that incorporates DNA and protein sequence data and bio-informatics resources acquired from the public domain and developed in-house. The foundation of PEDB is the identification and characterization of a prostate transcriptome, the intermediary between the genome and the proteome that represents that portion of the human genome actively used or transcribed in the prostate.

This proposal work was designed to extend the PEDB capabilities by accomplishing the following specific objectives:

1) assemble and annotate a working prostate transcriptome;
2) develop a suite of database tools to facilitate investigator-initiated database queries;
3) extend the prostate transcriptome in 3 dimensions: a) acquiring rare transcripts; b) assembling sequences representing full-length genes, and c) mapping the locations of interesting and novel prostate genes.
4) assemble a solid-phase nonredundant archive of prostate-derived cDNA clones for distribution to investigators and to the Image Consortium sites.

During the work on the project, we extended the original aims to include the development of a parallel resource for studies of the mouse prostate. The mouse has become a valuable model organism for studies of prostate carcinogenesis, and we reasoned that a resource to facilitate these studies would be extremely useful. This report also describes our result in the context of the mouse prostate expression database (mpedb).

BODY

The following summarizes the technical objectives for the proposal and the work accomplished during the entire project period.

Technical objective 1: *To assemble and annotate a working prostate transcriptome* (months 1-16).

- *Task 1: Install Phrap, d2-cluster, and CAP3 software and test on small, known genomic sequence assemblies (months 1-3).* Completed. Phrap is the selected assembly algorithm of choice.
- *Task 2: Assemble UniGene and prostate EST test sets. Compare with previous assemblies performed with CAP2. Manually review assembly discrepancies. Compare assemblies with UniGene and CGAP clusters (months 1-6).* Completed. Phrap proved to be the most robust and accurate assembly tool.
- *Task 3: Assemble and annotate all PEDB sequences using best available algorithm (months 6-12).* Completed. We have completed the download of 20,000 additional chromatograms from the Washington University web site and added an additional 5,000 ESTs from our own sequencing project. The assembly of these ESTs with

phrap has been completed and the contigs have been annotated against sequences housed in Genbank and Unigene. The latest assembly statistics are shown in Figure 1 with the assembly schema and results.

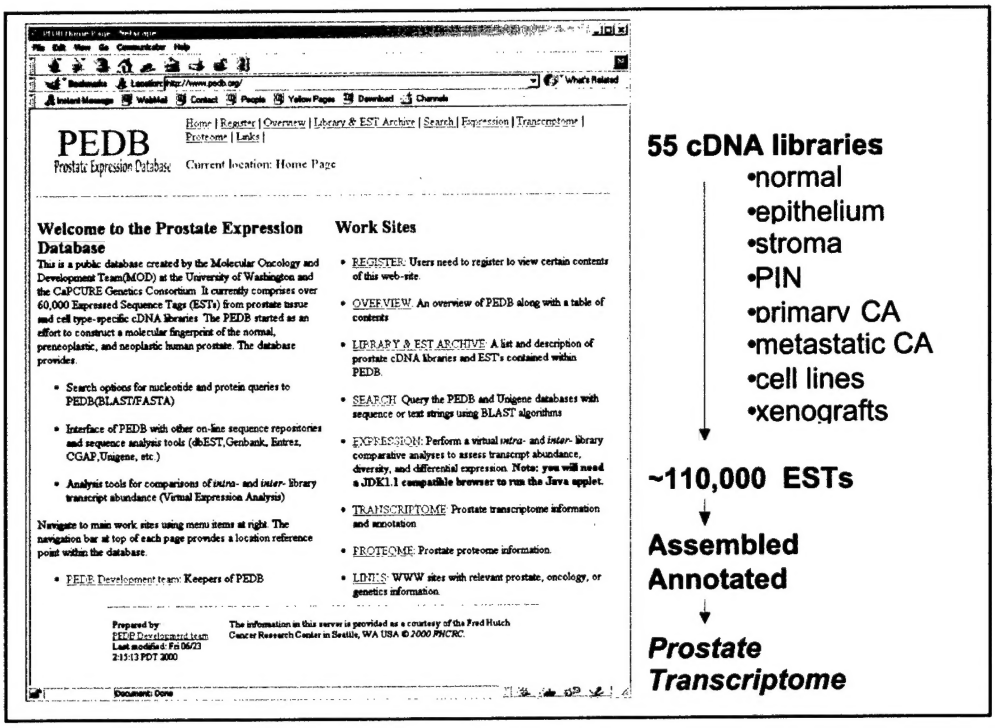


Figure 1. (above) WWW Interface for the Prostate Expression Database (PEDB). 55 libraries containing **~110,000 EST**s have been processed, assembled, and annotated to comprise a Prostate Transcriptome.

- *Task 4: Develop a gene classification schema based upon function, and automate assignments of clusters to functional groups (months 8-16).* We have completed a functional annotation scheme modeled after the TIGR annotation scheme for partitioning genes into cellular functional groups. This has been applied to the LNCaP dataset and is available for viewing and analysis on the PEDB website.
- *Task 5: Develop a Graphical User Interface for viewing and navigating between sequence, functional group, and expression data (months 3-12).* A user interface has been developed for the viewing of sequence chromatograms and for searching the database with keywords in addition to BLAST queries.
- *Task 6: Write scripts to automate the input of new prostate ESTs, processing of new ESTs, clustering of the database sequences and annotation of the entire cluster complement on a monthly basis (months 12-16).* Completed.

Technical objective 2: To develop a suite of database tools to facilitate investigator-initiated database queries (months 1-18). Completed. We developed a search

interface to incorporate searches of sequences with truncated key words, wildcards, and sequence-based queries using the basic local alignment search tool (BLAST).

- *Task 7: Evaluate potential sequence cluster/assembly viewing tools: DrawMap, Consed, Phrapview, CloneView, and AlignView (months 6-12).* We have completed the evaluation of sequence assembly viewing tools and have selected Consed as the viewer of choice.
- *Task 8: Design a client application (ContigView) extending the functionality of the Virtual Expression Analysis Tool to view the cluster output produced by the best algorithm as identified in specific aim 1 (months 8-14).* Completed. This objective was achieved by providing an option for down-loading the contig sequence, rather than actually viewing the cluster. We opted not to invest further programming time into this tool, as in the near future the code for such a viewing tool will be available through the ongoing viewing and annotation projects for viewing the human and mouse draft sequences.
- *Task 9: Design GUI to support high level viewing of clustered data with graphical maps incorporating zoom features for viewing nucleotide sequence traces and assemblies (months 8-14).* A tool for viewing individual sequence traces has been developed and implemented into PEDB. Consed is used for viewing assemblies.
- *Task 10: Write Java code for ContigView and test on datasets representing assemblies of few and many ESTs with both short and long consensus sequences (months 10-24).* We have opted not to incorporate Java code for this process. Current software was developed in C and PERL. Consensus sequences are available for down-load from the PEDB website. Sequences utilized for the consensus sequence assembly are listed adjacent to each PEDB species number and are also downloadable.
- *Task 11: Test (and modify if necessary) ContigView on Windows/NT/MacIntosh/Unix operating systems (months 24-30).* This objective has been met by providing the resource via the PEDB website.
- *Task 12: Write applications to link cluster consensus to relevant public databases (Genbank, etc) (months 20-24).* Completed. Sequences are referenced against Genbank, dbEST, and Unigene.
- *Task 13: Write applications for integrating gene analysis tools: exon prediction, promoter finders, transcription factor binding site ID, protein motif ID (months 18-28).* This objective is obsolete. The availability of browser functions at the SantaCruz human genome assembly site satisfies this objective and resources were not used 'reinventing' these applications (http://genome.ucsc.edu/).
- *Task 14: Evaluate the incorporation of software for SNP detection (PolyPhred) in client-selected PEDB clusters (months 26-30).* We have evaluated the utility of incorporating EST-derived SNPs. After discussions with Dr. Debbie Nickerson (University of Washington), the accuracy of EST-derived SNPs, and hence their utility for polymorphism detection, is suspect. The accuracy is such that re-sequencing is required for confirmations. Large-scale SNP discovery efforts obviate the need for further efforts on this task.

Technical objective 3: *To extend the prostate transcriptome in 3 dimensions: 1) acquire rare transcripts 2) assemble sequences representing full-length genes and 3) map the location to EST clusters to specific chromosomal sites (months 12-25).*

- *Task 15: construct LNCaP random primed library and CAP-finder library (months 6-7).* We have constructed one prostate cDNA library from androgen stimulated LNCaP and one cDNA library from androgen-starved LNCaP and compared these expression profiles to identify androgen-regulated genes in prostate epithelium (see Clegg et al #1 in reportable outcomes). We have constructed one prostate cDNA library from prostate small cell carcinoma and sequenced approximately 3,500 ESTs (see Clegg et al #2 in reportable outcomes. Virtual comparison's of this library against prostate adenocarcinoma and small cell lung carcinoma has identified several known genes and novel sequences that may be useful for studying the development, progression, and therapy of this variant of prostate cancer (see Clegg et al #2 in reportable outcomes).
- *Task 16: partially sequence 1,600 cDNAs from each library and enter ESTs into PEDB (months 8-12).* See above. Approximately 2,000 additional ESTs from the LNCaP libraries and 3,500 ESTs from the prostate small cell library have been entered into PEDB, assembled using phrap, and annotated against sequences present in the public nucleotide databases.
- *Task 17: as above with normal prostate tissue (months 13-18).* We have constructed cDNA libraries from microdissected luminal cell, basal cell, and stromal tissue. A total of 2,200 ESTs have now been produced from these libraries.
- *Task 18: as above with microdissected primary prostate cancer tissue (months 25-30).* Two libraries representing primary prostate carcinoma have now been constructed. Approximately 1,000 ESTs were generated per library.
- *Task 19: "Negative Select" 10,000 cDNAs from normal prostate cDNA array (months 19-20).* We have now selected ~6,500 non-redundant clones for array construction, and used these in the constuction of 1st and 2nd generation prostate microarrays.
- *Task 20: partially sequence 10,000 negatively selected, low abundance cDNAs and submit ESTs into PEDB (months 21-25).* Completed.
- *Task 21: Identify 60 interesting uncharacterized prostate ESTs/cDNAs based upon a) homology to known physiologically important genes or b) novelty, to directly obtain full-length cDNA sequence using RACE, library screening, genomic assembly, and primer-directed sequencing. A total of 15 full-length cDNAs per year will be obtained (ongoing throughout period of award).* Since the previous progress report, we have cloned and sequenced the full-length of 4 androgen-regulated prostate cDNAs; PWDMP, PART2, 6A4, and KIAA0056. Characterization of these genes is in progress. A report of the characterization of PWDMP has recently been published (see reportable outcomes Lin et al).
- *Task 22: Map interesting prostate cDNAs described above using radiation hybrid panel mapping (ongoing throughout period of award).* With the completion of the human genome project, this aim is essentially obsolete. We have now defined the chromosomal locations for >5,000 prostate ESTs using the available human genome sequence.
- *Task 23: submit data to PEDB and public databases.* Completed. EST sequences have been deposited in the public sequence repositories.

Technical objective 4: *To assemble a solid phase nonredundant archive of prostate-derived cDNA clones. (months 10-36). Completed.*

- *Task 24: identify a cohort consisting of 3,000 distinct, unique prostate clusters from year 1 PEDB assembly (month 10).* We have assembled a non-redundant set of 6,500 cDNAs from LNCaP, normal and neoplastic prostate cDNA libraries. These have now been re-arrayed into 96-well and 384-well microtiter plates. The clone set has been replicated. PCR amplification has been performed.
  * In addition to this effort, we have initiated an effort to assemble a working transcriptome of the murine prostate gland. To this end we have constructed 14 mouse prostate cDNA libraries, sequenced >50,000 ESTs, and assembled a non-redundant clone set of >10,000 cDNAs. Sequence verification was performed. This has been followed by the construction of a mouse prostate cDNA array.
- *Task 25: cross-reference cluster sequences with PEDB clone archive to determine the clones physically available for biological studies (month 11).* Completed.
- *Task 26 determine the longest physical clone for each cluster and consolidate bacterial transformants into 96-well plates using the Genetix Q-bot. Preserve for storage (months 12-13).* Completed.
- *Task 27: annotate and ship to IMAGE consortium clone distributors (month 14).* Clones have been replicated into shipping sets and are available for users desiring these clones.
- *Task 28: repeat Tasks 24-27 for 3,000 additional unique clusters at the end of month 24.* Completed.
- *Task 29: repeat Tasks 24-27 for 3,000 additional unique clusters at the end of month 35.* Completed.
- *Task 30: plan for incorporation and integration of PEDB with microarray data and proteomics data (months 24-36).* We have obtained database software for archiving and analyzing microarray data from Stanford University and the Institute for Systems Biology. We have entered microarray data into PEDB in formats suitable for downloading and analysis. Future work to develop direct analysis tools within PEDB is in the planning stages.
- *Task 31: analyze/compile data and prepare formal report (month 36).* Completed.

KEY RESEARCH ACCOMPLISHMENTS

- Selected phrap as the sequence assembly algorithm for PEDB. Assembled and annotated >110,000 PEDB ESTs. Manuscript published (Nelson et al-2002)
- Constructed cDNA libraries from microdissected luminal epithelial cells, basal epithelial cells, stromal cells, and primary prostate carcinoma.
- Sequenced cDNAs from LNCaP (androgen-stimulated and androgen-starved), luminal cell, basal cell and stromal cell cDNA libraries (total ~10,000 ESTs) and assembled the ESTs into clusters/contigs. The data indicate the libraries are of good quality with significant diversity.
- Virtual comparison of the LNCaP libraries identified 4 additional new androgen-regulated genes. Northern analysis confirmed androgen-regulation for these genes. Manuscripts published (Moore et al-2002, Lin et al-2003)
- Constructed a cDNA library of prostate small cell carcinoma. 3,000 cDNA clones have been sequenced and deposited into PEDB. Manuscript published (Clegg et al-2003).

- Compiled a non-redundant virtual and physical archive of prostate ESTs/cDNAs comprising 6,500 distinct species. These clones have been consolidated, replicated, and arrayed for cDNA microarray analysis. This resource has been used in several gene expression studies. Manuscripts published (Nelson et al-2002; Bonham et al-2002).
- Initiated the characterization of the mouse prostate transcriptome. Constructed 14 mouse prostate cDNA libraries and produced >50,000 ESTs. This resource has now been shared with 6 other investigative teams studying prostate carcinoma (Pradip Roy-Burman-USC; Hong Wu-UCLA; Robert Sikes-University of Delaware; Cory Abate-Shen-RWJ New Jersey; Leland Chung-Emory; Robert Vessella-University of Washington).

REPORTABLE OUTCOMES

Clegg N, Eroglu B, Ferguson C, Arnold H, Moorman A, **Nelson PS.** (2002) Digital expression profiles of the prostate androgen-response program. *J Steroid Biochem Mol Biol* 80(1):13-23.

Liu AY, **Nelson PS**, van den Engh G, Hood L. (2002) Human prostate epithelial cell-type cDNA libraries and prostate expression patterns. *Prostate* 50(2):92-103.

Clegg N, Ferguson C, True LD, Arnold H, Moorman A, Quinn, JE, Vessella RL, and **Nelson PS**. (2003) Molecular characterization of Prostatic small-cell neuroendocrine carcinoma. *Prostate* 55(1):55-64.

Nelson PS, Pritchard C, Abbott D, Clegg N. (2002) The human (PEDB) and mouse (mPEDB) Prostate Expression Databases. *Nucleic Acids Res* 30(1): 218-220.

Bonham M, Arnold H, Montgomery B, **Nelson PS**. (2002) Molecular effects of the herbal compound PC-SPES: identification of activity pathways in prostate carcinoma. *Cancer Research* 62(14):3920-3924.

Moore S, Pritchard C, Lin B, Ferguson C, **Nelson PS.** (2002) Isolation and characterization of the murine prostate short-chain dehydrogenase/reductase 1 (Psdr1) gene, a new member of the short-chain steroid dehydrogenase/reductase family. *Gene* 293(1-2):149-160.

Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, White J, Hood L, Lin B. (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci USA* 99(18):11890-11895.

Lin B, White JT, Utleg AG, Wang S, Ferguson C, True LD, Vessella R, Hood L, **Nelson PS.** (2003) Isolation and characterization of human and mouse WDR19, a novel WD-repeat protein exhibiting androgen-regulated expression in prostate epithelium. *Genomics*-in press.

CONCLUSIONS

The research accomplished through the specific aims of this proposal has assembled a working virtual prostate transcriptome that defines the genes used or transcribed in prostate cell types and tissues. This transcriptome has a physical counterpart of >10,000 cDNAs arrayed in cDNA microarray format for large-scale expression studies. This transcriptome has been used as a foundation for studies of the prostate proteome, the working counterpart to the genome and

transcriptome. Our results show that these approaches are complementary. Analysis of the virtual transcriptome of LNCaP cells has identified 15 new androgen-regulated genes to date. The characterization of several of these genes have been published, and the characterization of others are in progress. We have extended the human PEDB to also encompass murine prostate gene expression in mPEDB. mPEDB has already been proven to be a valuable resource for research involving murine models of prostate carcinoma. Immediate uses for mPEDB involve comparative gene expression studies with the human prostate.

The importance of the research accomplishments center on 1) providing a virtual and physical set of prostate genes that can be used by investigators in assessing changes in gene expression that associate with prostate carcinogenesis. These genes may serve as diagnostic or therapeutic targets; 2) providing a resource for thoroughly assessing mouse models of prostate carcinoma. This assessment can be used in identifying cancer-related genes as well as assessing responses to therapy in pre-clinical trials.

REFERENCES

None.

LIST OF PERSONNEL

Denise Abbott
Nigel Clegg
Burak Eroglu
Camari Ferguson
David Gifford
Cynthia Heinlein
Masazumi Matsumura
Peter Nelson
Heinrich Renneberg
Anupama Vasanth
Gabrielle Zecha

BIBLIOGRAPHY/APPENDICES

Clegg N, Eroglu B, Ferguson C, Arnold H, Moorman A, **Nelson PS**. (2002) Digital expression profiles of the prostate androgen-response program. *J Steroid Biochem Mol Biol* 80(1):13-23.

Liu AY, **Nelson PS**, van den Engh G, Hood L. (2002) Human prostate epithelial cell-type cDNA libraries and prostate expression patterns. *Prostate* 50(2):92-103.

Clegg N, Ferguson C, True LD, Arnold H, Moorman A, Quinn, JE, Vessella RL, and **Nelson PS**. (2003) Molecular characterization of Prostatic small-cell neuroendocrine carcinoma. *Prostate* 55(1):55-64.

Nelson PS, Pritchard C, Abbott D, Clegg N. (2002) The human (PEDB) and mouse (mPEDB) Prostate Expression Databases. *Nucleic Acids Res* 30(1): 218-220.

Bonham M, Arnold H, Montgomery B, **Nelson PS**. (2002) Molecular effects of the herbal compound PC-SPES: identification of activity pathways in prostate carcinoma. *Cancer Research* 62(14):3920-3924.

Moore S, Pritchard C, Lin B, Ferguson C, **Nelson PS.** (2002) Isolation and characterization of the murine prostate short-chain dehydrogenase/reductase 1 (Psdr1) gene, a new member of the short-chain steroid dehydrogenase/reductase family. *Gene* 293(1-2):149-160.

Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, White J, Hood L, Lin B. (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci USA* 99(18):11890-11895.

Lin B, White JT, Utleg AG, Wang S, Ferguson C, True LD, Vessella R, Hood L, **Nelson PS.** (2003) Isolation and characterization of human and mouse WDR19, a novel WD-repeat protein exhibiting androgen-regulated expression in prostate epithelium. *Genomics* - in press.

PERGAMON

Journal of Steroid Biochemistry & Molecular Biology 80 (2002) 13–23

The Journal of
Steroid Biochemistry
&
Molecular Biology

www.elsevier.com/locate/jsbmb

Digital expression profiles of the prostate androgen-response program

Nigel Clegg, Burak Eroglu, Camari Ferguson, Hugh Arnold, Alec Moorman, Peter S. Nelson*

Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA

Received 15 June 2001; accepted 24 September 2001

Abstract

The androgen receptor (AR) and cognate ligands regulate vital aspects of prostate cellular growth and function including proliferation, differentiation, apoptosis, lipid metabolism, and secretory action. In addition, the AR pathway also influences pathological processes of the prostate such as benign prostatic hypertrophy and prostate carcinogenesis. The pivotal role of androgens and the AR in prostate biology prompted this study with the objective of identifying molecular mediators of androgen action. Our approach was designed to compare transcriptomes of the LNCaP prostate cancer cell line under conditions of androgen depletion and androgen stimulation by generating and comparing collections of expressed sequence tags (ESTs). A total of 4400 ESTs were produced from LNCaP cDNA libraries and these ESTs assembled into 2486 distinct transcripts. Rigorous statistical analysis of the expression profiles indicated that 17 genes exhibited a high probability ($P > 0.9$) of androgen-regulated expression. Northern analysis confirmed that the expression of *KLK3/PSA*, *FKBP5*, *KRT18*, *DKFZP564K247*, *DDX15*, and *HSP90* is regulated by androgen exposure. Of these, only *KLK3/PSA* is known to be androgen-regulated while the other genes represent new members of the androgen-response program in prostate epithelium. LNCaP gene expression profiles defined by two independent experiments using the serial analysis of gene expression (SAGE) method were compared with the EST profiles. Distinctly different expression patterns were produced from each dataset. These results are indicative of the sensitivity of the methods to experimental conditions and demonstrate the power and the statistical limitations of digital expression analyses.

Keywords: Androgen; Prostate; EST; SAGE; Transcriptome

1. Introduction

Genes regulated by androgenic hormones are of critical importance for the normal physiological function of the human prostate gland, and they contribute to the development of prostate diseases such as benign prostatic hypertrophy (BPH) and prostate carcinoma. Androgens such as testosterone and dihydrotestosterone (DHT) interact with the androgen receptor (AR) leading to the transcriptional activation of androgen-target genes [1]. This gene network regulates prostate morphogenesis, growth, and function, and promotes the development and progression of prostate neoplasia [2]. Despite the importance of androgens in modulating diverse prostate cellular processes, relatively few components of this androgen-response program have been identified or characterized.

Current estimates indicate that between 35,000 and 40,000 genes are encoded in the human genome [3,4]. To confer developmental and functional specificity, only a fraction of this total is transcribed in a given tissue or cell type at any given time. This repertoire of expressed genes in transcript form is termed the transcriptome [5], a dynamic assessment or inventory of gene expression activity that reflects the cellular developmental state and response(s) to environmental perturbations. Proceeding from the hypothesis that comprehensive gene expression profiles will provide insights into cellular function, several procedures have been developed to qualitatively and quantitatively assess transcriptomes. These methods can be broadly divided into analog approaches

Abbreviations: *KLK3*, kallikrein 3; *RPLP0*, ribosomal protein large, P0; *UQCRC2*, ubiquinol-cytochrome *c* reductase core protein 2; *FKBP5*, FK506-binding protein 5; *DKFZP564K247*, DKFZP564K247 protein; *PHGDH*, phosphoglycerate dehydrogenase; *KRT18*, keratin 18; *RPS25*, ribosomal protein S25; *EIF3S6*, eukaryotic translation initiation factor 3, subunit 6 (48 kDa); *FTL*, ferritin, light polypeptide; *DDX15*, DEAD/H (Asp-Glu-Ala/His) box polypeptide; *RPS27A*, ribosomal protein S27A; *ACADVL*, acyl-coenzyme A dehydrogenase, very long chain; *KIAA0101*, KIA0101 gene product; *DKFZP564D0462*, hypothetical protein DKFZP-564D0462; *RPS15A*, ribosomal protein S15a; *DED*, apoptosis antagonizing transcription factor; *BSG*, basigin; *TPI1*, triosephosphate isomerase 1; *CLTB*, clathrin, light polypeptide (Lcb); *DBI*, diazepam binding inhibitor; *ENO1*, enolase 1 (alpha); *KLK2*, kallikrein 2; *KLK4*, kallikrein 4; *ODC1*, ornithine decarboxylase 1; *PDHA1*, pyruvate dehydrogenase (lipoamide) alpha 1; *TMEPA1*, transmembrane, prostate androgen-induced RNA; *TUBA1*, tubulin, alpha 1; *UGT2B17*, UDP glycosyltransferase 2 family, polypeptide B17; *VEGF*, vascular endothelial growth factor

* Corresponding author. Fax: +1-206-667-2917.
E-mail address: pnelson@fhcrc.org (P.S. Nelson).

0960-0760/02/$ – see front matter
PII: S0960-0760(01)00167-4

such as DNA array analysis [6–8], and digital methods as exemplified by expressed sequence tag (EST) quantitation [9] and the serial analysis of gene expression (SAGE) [10]. Each approach has distinct advantages and limitations that have been detailed previously [11]. A principle advantage of digital methods is the possibility of sampling the complete transcriptome in a single experiment. These approaches also permit the analysis of previously uncharacterized genes and allow for direct statistical analyses of transcript numbers rather than relying on indirect measures of transcript ratios.

Our objective in this study was to identify genes expressed in human prostate cells exhibiting transcriptional regulation by androgens. We hypothesize that such genes could be direct mediators of the androgen-receptor pathway or be involved in prostate-specific functions that could be exploited for understanding normal and neoplastic prostate growth. To facilitate systematic studies of prostate gene expression, we have established the prostate expression database (PEDB), an archive that contains more than 70,000 ESTs generated from prostate cDNA libraries [12]. Two libraries constructed specifically for this study comprise genes expressed in the LNCaP prostate cancer cell line under conditions of androgen stimulation and androgen deprivation. The LNCaP cell line represents a model system for the study of androgen regulation as LNCaP cells express a functional AR, proliferate in response to physiological levels of androgens, and increase the transcription of known androgen-regulated genes such as prostate specific antigen (PSA) [13]. We applied statistical tools to compare these EST datasets and identified both known and novel genes with a high probability ($P > 0.9$) of being regulated by androgens. Northern analysis was used to confirm androgen-regulated expression. These studies identified *FKBP5*, *KRT18*, *DKFZP564K247*, *DDX15*, and *HSP90*, as new members of the prostate epithelial androgen-response program. LNCaP transcriptomes defined by two distinct SAGE experiments were also examined for genes exhibiting androgen regulation and these results were compared with the EST profiles. These results support the use of comprehensive gene expression profiling methods to define cellular responses to hormonal stimuli, and demonstrate both the power and the statistical limitations of digital expression analyses.

2. Materials and methods

2.1. Cell culture

The prostate carcinoma cell line LNCaP was obtained from ATCC and grown in RPMI 1640 with 10% FCS (Life Technologies, Inc.). Cells were transferred into RPMI-1640 medium with 10% charcoal-stripped fetal calf serum (CS-FCS) 24 h before androgen-regulation experiments. This medium was replaced with fresh CS-FCS media or fresh CS-FCS including 1 nM of the synthetic androgen R1881 (New England Nuclear Life Science Products, Inc.). Cells were harvested for RNA isolation at 0- and 24-h time points.

2.2. Library construction

Total RNA was isolated from androgen-stimulated (LNCaP01) and androgen-starved (LNCaP02) cells using TRIzol (Life Technologies, Inc.) according to the manufacturer's instructions. Poly(A)$^+$ RNA was purified using oligo(dT) chromatography [14]. A unidirectional library was constructed in the pSport1 vector (Life Technologies, Inc.) according to a modification of the Gubler and Hoffman [15] protocol. Poly(A$^+$) was reverse-transcribed using superscript reverse transcriptase and an oligo(dT) linker/primer containing a *Not*1 site (Life Technologies). Sephacryl-S400 (Pharmacia) was used to size-select the synthesized cDNA and remove excess linkers. Blunt-ended, double-stranded cDNA was ligated with a *Sal*1 adapter, digested with *Not*1, then ligated into *Sal*1–*Not*1 digested pSport1. High-efficiency electrocompetent *Escherichia coli* were transformed using a Bio-Rad GenePulser under recommended conditions. Approximately, 86% of the LNCaP01 and 89% of the LNCaP02 transformants contained inserts. The average insert size for the library was 1.7 kb.

2.3. DNA sequencing

Independent transformant colonies were picked into 100 ul PCR mix [10 mM Tris, pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl, 120 uM dNTPs, 1 U Taq polymerase (Promega) and 0.12 uM each of VN26 TTTCCCAGTCACGACGTTGTA and VN27 GTGAGCGGATAACAATTTCAC] and subjected to 40 cycles of 30 s at 94 °C, 30 s at 60 °C and 120 s at 72 °C followed by 10 min at 72° C. Amplified inserts were purified over Sephacryl S-500 (Pharmacia), and 4 ul was used in DNA sequencing reactions using M13 reverse fluorescent-labeled dye primers as detailed in the Prism cycle sequencing kit (Applied Bio-systems, Inc.). Reaction products were electrophoresed on ABI 373 and 377 DNA sequencers.

2.4. Northern analysis

Total RNA was isolated from LNCaP cells using the TRIzol method according to the manufacturer's instructions. Ten micrograms of total RNA was fractionated on 1.2% agarose gels under denaturing conditions and transferred to nylon membrane using the capillary method. Blots were hybridized with cDNA probes labeled with [α-^{32}P]-dCTP using a Random Primers DNA labeling kit (Life Technologies Inc.) according to the manufacturer's protocol. Filters were imaged and quantitated using a phosphor-capture screen and Image Quant software (Molecular Dynamics). β-Actin was used as an internal control for normalizing transcript levels between samples.

2.5. EST assembly, annotation, and comparison

DNA sequences were stored, clustered, and annotated using the PEDB relational database management tools and data analysis pipeline [17].[1] Briefly, vector, *E. coli*, and interspersed repeats were masked in the ESTs using Cross_Match[2] and Repeatmasker.[3] Poor quality sequences, with >50% ambiguous nucleotides ('N') between nucleotides 100 and 500 were discarded. CAP2 [16], a multiple sequence alignment program based on a variant of the Smith–Waterman algorithm, was used to cluster the masked sequence and generate a consensus sequence for each assembly. Each distinct cluster was annotated by searching Unigene,[4] GenBank,[5] and dbEST[6] databases using BLASTN.[7] Annotations were assigned automatically using SmartBlast (Perl 5.0) to select the database match with the lowest *P*-value and the highest BLAST score where the maximum *P*-value was e^{-20} and the minimum BLAST score was 500. Some species required manual reconciliation when either two distinct PEDB species were annotated with the same identification, or when annotations differed between public databases. The Virtual Expression Analysis Tool (VEAT[8]) and scripts written in Perl 5.0 were used for creating transcript species reports. The biological role for each species was assigned using the categories described by Adams et al. [9]. Supplemental information, including a complete list of species and transcript frequencies is available at the PEDB web site. Gene symbols are from the HUGO Gene Nomenclature Committee.

Using statistics described by Audic and Claverie [11], differential gene expression in androgen-stimulated and androgen-deprived cells was inferred based on differential representation of ESTs in cDNA libraries.

2.6. SAGE data acquisition and analysis

The following LNCaP SAGE libraries are listed at the NCBI Library Browser web site[9] and were downloaded from SAGE-map's anonymous FTP site[10]: SAGE_Chen_LNCaP (62,681 tags), SAGE_Chen_LNCaP_no-DHT (65,206 tags), SAGE_CPDR_LNCaP-C (41,848 tags), and SAGE_CPDR_LNCaP-T (44,370 tags). For simplicity, these libraries are hereafter called LNCaP(+)DHT, LNCaP(−)DHT, LNCaP-C and LNCaP-T. Statistical analyses were performed using the software provided at the SAGEmap xProfiler web site.[11]

[1] http://www.pedb.org.
[2] http://www.genome.washington.edu/UWGC/methods.htm.
[3] http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker.
[4] ftp://ncbi.nlm.nih.gov/repository/UniGene/Hs.seq.all.Z.
[5] ftp://ncbi.nlm.nih.gov/blast/db/nt.Z.
[6] ftp://ncbi.nlm.nih.gov/blast/db/est.Z.
[7] http://blast.wustl.edu.
[8] http://www.pedb.org.
[9] http://www.ncbi.nlm.nih.gov/SAGE/sagelb.cgi.
[10] ftp://ncbi.nlm.nih.gov/pub/sage/seq/.
[11] http://www.ncbi.nlm.nih.gov/SAGE/sageexpsetup.cgi.

3. Results

3.1. EST-derived LNCaP transcriptomes

Two cDNA libraries, LNCaP01 and LNCaP02, were constructed from the prostate adenocarcinoma cell line LNCaP under conditions of androgen stimulation and androgen starvation, respectively. Approximately, 2300 ESTs were produced from each library and the sequences were entered into the PEDB [12]. Automated processing of the ESTs to remove short, poor quality, repetitive, and/or vector sequences eliminated 779 ESTs from further analysis. The remaining 4458 ESTs were assembled using the CAP2 sequence assembly program. Each EST cluster was annotated by searching the Unigene, GenBank, and dbEST databases with the CAP2-generated cluster consensus sequences using BLASTN. Clusters annotated with the same database sequence were joined, and all ESTs grouped to the same cluster were assigned the same unique PEDB cluster ID. ESTs for mitochondrial genes were grouped as a single cluster and accounted for approximately 6% of all ESTs. These genes were not further analyzed. In total, 2486 distinct transcript species were identified (Fig. 1): 2240 were homologous to previously identified genes or ESTs, and 252 were not significantly homologous to any public database sequence. The latter species may represent novel genes or previously unsequenced regions of known genes.

The number of distinct transcripts comprising the LNCaP01 and LNCaP02 transcriptomes are quantitatively similar, but qualitatively different. In all, 87% of the species were represented in one transcriptome or the other, but not in both (Fig. 1A). Despite the difference in species composition, the EST frequency distributions of the two samples were similar: nearly 78% of the species are represented by a single EST and only 9% were composed of more than 2 ESTs (Table 1). These distributions are broadly consistent with previous estimates which indicate there are relatively few transcripts expressed in high abundance (5–15 species at 10,000 copies per cell), an intermediate number of moderately abundant transcripts (500 species at 300 copies per cell) and many low abundance transcripts (10,000 different species expressed in 1–15 copies per cell) [17]. In all, 70% of the transcript species with two or more ESTs in either LNCaP01 or LNCaP02 were also present in the other library (Fig. 1B). Thus, while few low abundance transcripts were found in both datasets, most of the high abundance transcripts were found in common.

Functional roles were assigned to each distinct species according to the convention established by Adams et al. [9]. The five primary biological roles were cell division, cell signaling/cell communication, cell structure/motility, cell/organism defense, and metabolism. For graphical presentation, we added the 'androgen-regulated' category to emphasize the primary difference between the experimental samples (Fig. 2). In total, 923 transcript species could be

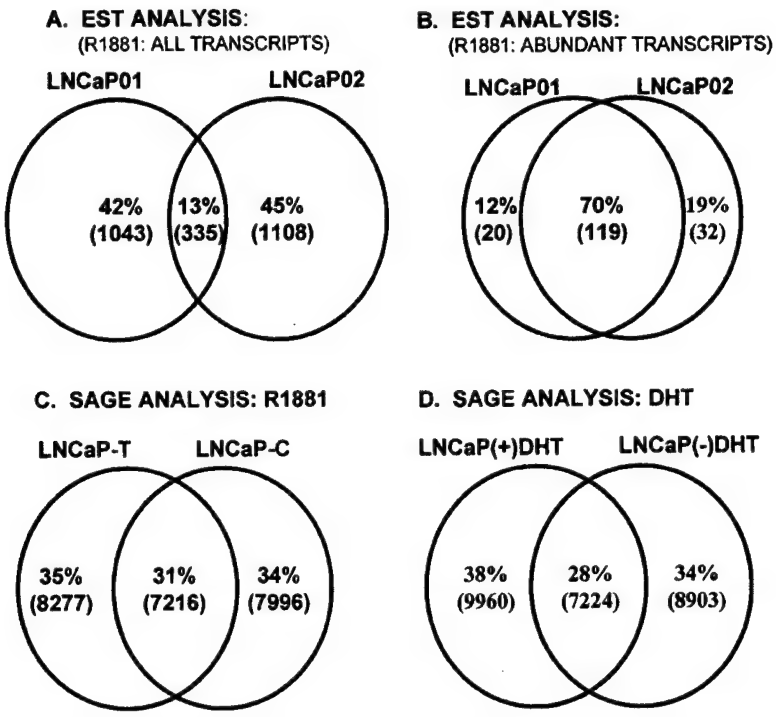


Fig. 1. Summary of LNCaP transcriptome diversity determined by EST and SAGE analysis. Representations of (A) the EST-derived number of all distinct transcripts unique to two LNCaP cell states (synthetic androgen R1881-stimulated LNCaP, LNCaP01; and R1881-starved LNCaP, LNCaP02) and those expressed in common between the two cell states; (B) the EST-derived number of highly and moderately expressed transcripts in LNCaP01 and LNCaP02 (>2 ESTs in one or both libraries) and those expressed in common; (C) SAGE analysis determining the number of distinct transcripts unique and in common between R1881-stimulated and starved LNCaP cells; (D) SAGE analysis determining the number of distinct transcripts unique and in common between DHT-stimulated and starved LNCaP cells.

Table 1
Distribution of molecular species by EST frequency

ESTs/species	No. of species (proportion of total)	
	LNCaP01	LNCaP02
1	1064 (0.78)	1133 (0.79)
2	202 (0.15)	199 (0.14)
3	55 (0.04)	56 (0.04)
4	26 (0.02)	23 (0.02)
5	8 (0.01)	8 (0.01)
6	6 (<0.01)	8 (0.01)
>6	17 (0.01)	16 (0.01)
Total	1378	1443

assigned biological roles. A detailed annotation of LNCaP transcripts assigned to these functional roles can be viewed at the PEDB website.[12] Both LNCaP transcript profiles have a similar distribution of species in each functional category (Fig. 2). The protein/gene expression category is the largest, primarily because of the high frequency of ESTs for ribosomal proteins and translation factors. Similar results have been obtained for whole normal prostate tissue [18]. A comparison of the composition of broad functional categories does not reveal a cohort of genes that reflect androgen stimulation or starvation, but differential gene expression in response to androgens is clearly evident for individual genes (Fig. 2). *KLK3/PSA*, an androgen-regulated gene, represents 1.4% of the ESTs in LNCaP01 (derived from androgen-stimulated cells), but only 0.05% of the ESTs in LNCaP02. ESTs for the androgen-response genes *KLK2*, *KLK4*, *ODC1*, *TUBA1*, and *ENO1* were also more abundant in the LNCaP01 library.

3.2. Androgen-regulated genes identified by digital expression analysis

We compared the abundance of each transcript species represented in the androgen-stimulated and androgen-starved transcriptomes using a VEAT [12]. VEAT provides a comprehensive graphical view of transcript frequency, as defined by EST number, between two or more transcriptomes of interest (Fig. 3). Among the species with more than two ESTs in either library, the most extreme difference in EST frequency was observed for *KLK3/PSA*. Twenty-nine *KLK3/PSA* ESTs were isolated from LNCaP01, the library made from androgen-stimulated cells, and only one EST was isolated from LNCaP02 (Table 2). This finding was expected as *KLK3/PSA* is one of the most abundant transcripts

[12] www.pedb.org.

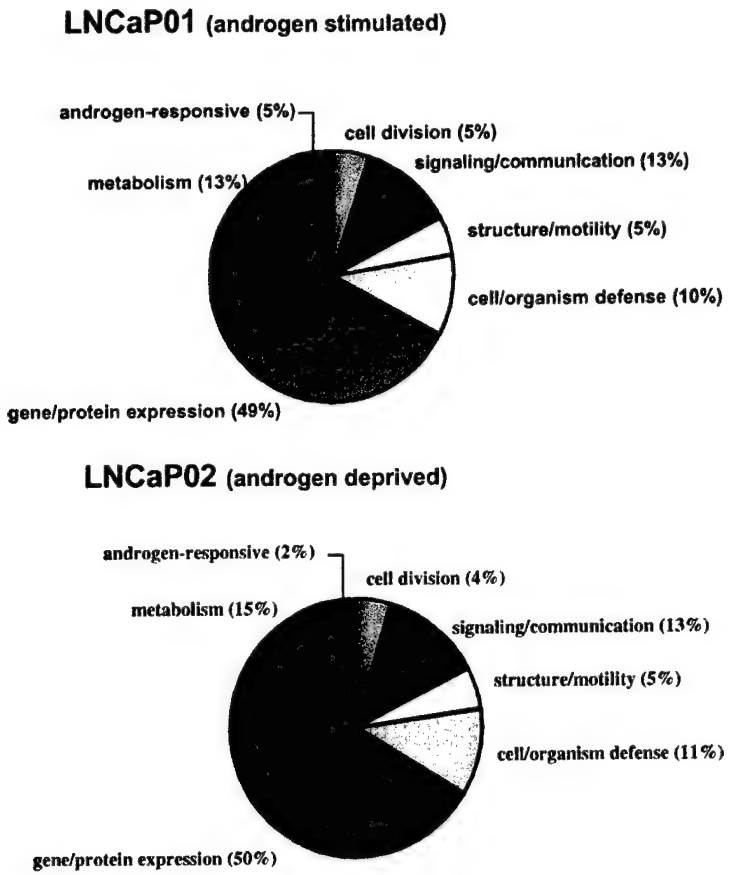


Fig. 2. Functional categorization of the LNCaP cell transcriptome. EST assemblies were annotated against the Genbank and Unigene databases. A putative functional role was assigned based upon categories developed by TIGR (http:www.tigr.org) and the percentage of ESTs corresponding to each role are depicted under cellular conditions of androgen stimulation and androgen starvation.

in the prostate [18] and is known to be transcriptionally regulated by androgens in LNCaP cells.

Additional differences in EST frequencies were seen for many other LNCaP transcripts. Determining the significance of these observations is challenging because of the potential for chance events (e.g. randomly selecting a given cDNA clone from a library) when the event is part of a large population of observable outcomes (e.g. cDNA libraries are complex and comprised of millions of cDNA clones). In order to validate and prioritize more subtle differences in gene expression, we used a statistical approach designed to provide a confidence interval indicating the probability that a given set of observations could occur by chance, or alternatively represents a significant change in expression [11]. Software available on the Internet [13] computes the confidence intervals corresponding to arbitrary significance levels and sample sizes of two datasets N_1 and N_2 [11]. Twenty-one species were predicted to be differentially expressed with a probability exceeding 90%: 9 were increased in response to androgens, and 12 were increased by androgen starvation (Table 2). With the exception of *KLK3/PSA*, none of these genes has previously been reported to be androgen-regulated in the prostate.

To confirm the differential expression statistics, the levels of transcription of *KLK3/PSA* and nine additional genes were examined by Northern analysis (Table 2, Fig. 4). cDNAs representing five different transcripts predicted to be androgen-upregulated by EST analysis were hybridized to Northern blots of RNA extracted from androgen-starved and androgen-stimulated LNCaP cells. Transcipts from each of the five genes were more abundant in androgen-stimulated cells than in androgen-deprived cells. Consistent with the EST frequency data, *KLK3/PSA* expression was increased 35-fold in androgen-stimulated cells compared to androgen-starved cells (Fig. 4). The transcripts encoding keratin 18 (*KRT18*), a gene expressed in prostate secretory cells, were increased 5-fold. FK506 binding protein 5 (*FKBP5*), *DKFZP564K247*, and *UOCRC2* were induced to a lesser extent. In contrast, statistical predictions were inaccurate for four of five putatively down-regulated genes. The steady-state level of DKFZp564K247 RNA was actually increased by androgens, and reduced transcription of

[13] http://igs-server.cnrs-mrs.fr.

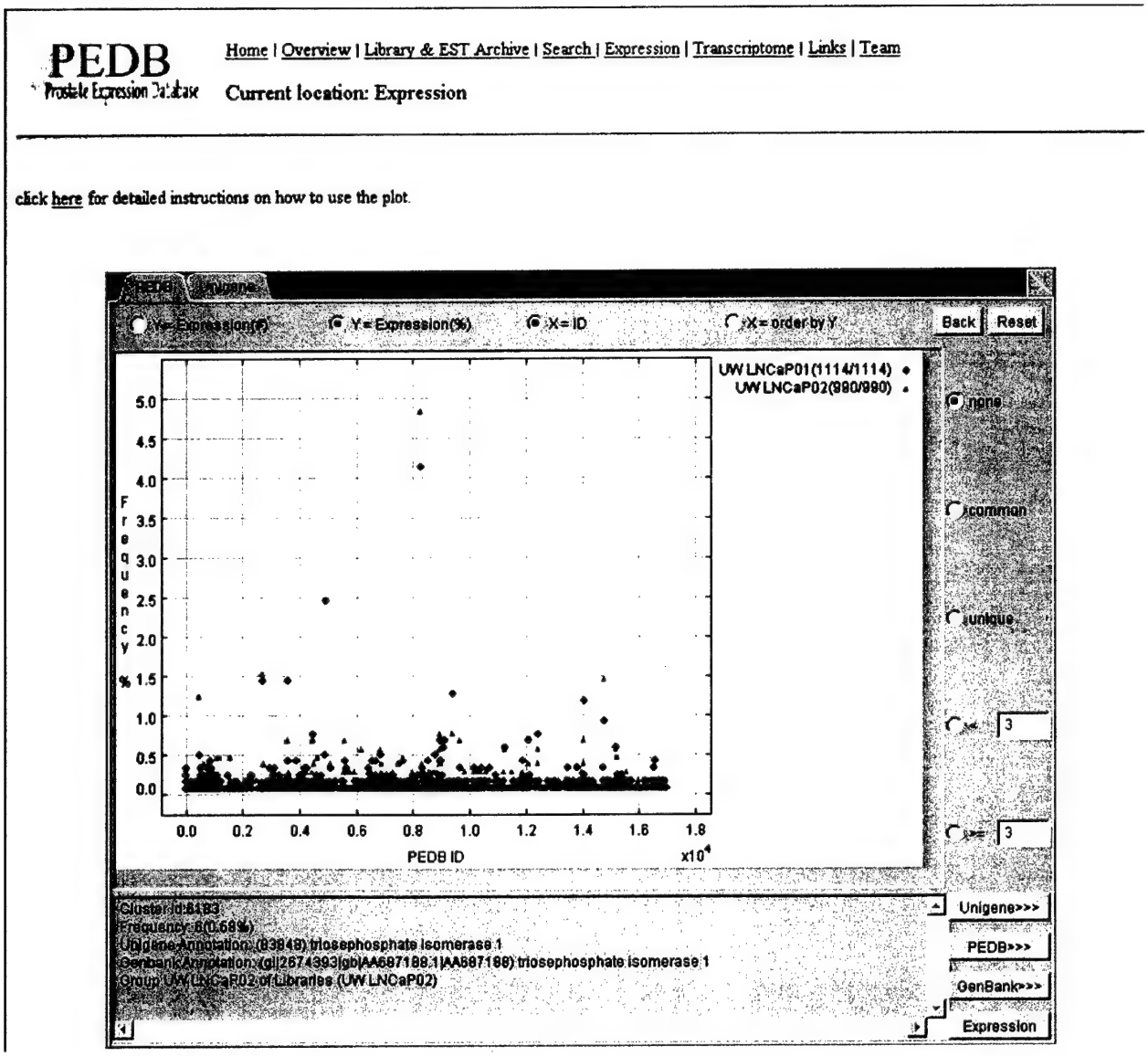


Fig. 3. Virtual differential expression determined by digital expression profiles. A view of cellular gene expression using the VEAT from the PEDB. Distinct transcripts are assigned a unique database ID and ordered along the *X*-axis. The number of ESTs assembled into each unique transcript (frequency) is displayed on the *Y*-axis as a percentage of the total EST number obtained from each library. Each library is represented by a different symbol (e.g. LNCaP01, triangle; LNCaP02, diamond). Highlighting any data point (using a mouse) provides annotation corresponding to that particular transcript (PEDB reference).

eukaryotic initiation factor 3 subunit 6 (*EIF3S6*), ribosomal protein 27a (*RPS27A*), and basigin (*BSG*) was not confirmed by Northern analysis. Surprisingly, one gene predicted to be decreased by androgen deprivation, the RNA helicase DEAD/H box polypeptide 15 (*DDX15*), was upregulated more than 3-fold by Northern analysis. There are several RNA helicases and our probe may be cross-hybridizing with another closely related androgen-inducible gene. At least, one other androgen-regulated RNA helicase has been reported [19].

In addition to the six androgen-responsive genes identified above, a heat shock protein gene (*HSP90*) was initially identified as androgen-regulated after a preliminary statistical analysis of approximately 1500 LNCaP01 and LNCaP02 ESTs. As the number of ESTs increased, *HSP90* was not differentially expressed based on the arbitrary statistical probability cut-off of $P > 0.90$; however, Northern blot analysis demonstrated a 4-fold increase in *HSP90* expression with androgen stimulation. There are numerous genes in the heat shock protein 90 gene family with strong sequence similarity [20], and our Northern hybridization conditions cannot differentiate between them. Nevertheless, this result confirms that one or more members of the *HSP90* gene family are androgen-responsive.

Table 2
Putative androgen regulated genes in LNCaP01/LNCaP02 libraries ($P \geq 0.9$) and corresponding SAGE data

Gene	ESTs			Androgen Response on Northern blot[a]	SAGE		
	No. of ESTs		Probability of differential expression[b]		SAGE Tag[c]	Probability of differential expression[d,e]	
	LNCaP01[f]	LNCaP02[g]				LNCaP-T/-C[h]	LNCaP(+)DHT/(−)DHT[i]
KLK3/PSA	29	1	$P > 0.99$	+35	GGATGGGGAT	$P = 1.00$ (82/5)	$P = 0.25$ (63/36)
RPLP0	22	9	$0.98 < P < 0.99$	nd[j]	CTCAACATCT	$P = 0.00$ (120/105)	$P = 0.00$ (248/292)
UQCRC2	5	0	$0.96 < P < 0.97$	+1.3	AAAGTCAGAA	$P = 0.16$ (6/8)	$P = 0.16$ (6/5)
FKBP5	4	0	$0.93 < P < 0.94$	+1.9	GTTCCAGTGA	$P = 0.66$ (6/0)	$P = 0.39$ (0/2)
DKFZP564K247	4	0	$0.93 < P < 0.94$	+1.7	TATCGGGAAT	–	$P = 0.29$ (2/1)
PHGDH	4	0	$0.93 < P < 0.94$	nd	TTACCTCCTT	$P = 0.22$ (22/12)	$P = 0.15$ (65/40)
KRT18	4	0	$0.93 < P < 0.94$	+5.0	CAAACCATCC	$P = 0.12$ (22/14)	$P = 0.02$ (27/35)
RPS25	6	1	$0.93 < P < 0.94$	nd	AATAGGTCCA	$P = 0.00$ (53/51)	$P = 0.06$ (132/84)
SFTPD	9	3	$0.90 < P < 0.91$	nd	–	–	–
EIF3S6	0	6	$0.98 < P < 0.99$	+1.2	AATATTGAGA	$P = 0.07$ (11/10)	$P = 0.33$ (12/6)
FTL	0	5	$0.96 < P < 0.97$	nd	CCCTGGGTTC	$P = 0.24$ (9/15)	$P = 0.15$ (22/37)
DDX15	0	4	$0.93 < P < 0.94$	+3.5	ATCGTTGTAA	$P = 0.37$ (4/1)	$P = 0.47$ (3/0)
RPS27A	0	4	$0.93 < P < 0.94$	+1.3	AACTAACAAA	$P = 0.15$ (16/10)	$P = 0.14$ (49/31)
ACADVL	0	4	$0.93 < P < 0.94$	nd	GCCGCCCTGC	$P = 0.13$ (6/6)	$P = 0.48$ (8/20)
KIAA0101	0	4	$0.93 < P < 0.94$	nd	ATGATTTATT	$P = 0.21$ (3/4)	$P = 0.47$ (3/0)
DKFZp564D0462	0	4	$0.93 < P < 0.94$	−2.6	CAGTTCTCAC	$P = 0.29$ (1/1)	$P = 0.40$ (2/0)
RPS15A	0	4	$0.93 < P < 0.94$	nd	GACAAAAAAA	$P = 0.26$ (27/14)	$P = 0.18$ (12/8)
RPS15A	–	–	–	–	GACTCTGGTG	$P = 0.16$ (11/7)	$P = 0.00$ (36/41)
DED	0	4	$0.93 < P < 0.94$	nd	GCACCTATTG	$P = 0.29$ (2/1)	$P = 0.35$ (0/1)
Species1145	0	4	$0.93 < P < 0.94$	nd	–	–	–
BSG	1	6	$0.92 < P < 0.93$	−1.02	GCCGGGTGGG	$P = 0.06$ (11/11)	$P = 0.00$ (216/341)
TPI1	1	6	$0.92 < P < 0.93$	nd	TGAGGGAATA	$P = 0.01$ (33/29)	$P = 0.02$ (39/32)

[a] Ratio of normalized signal intensity from RNA of hormone stimulated/starved cells.
[b] [11].
[c] Most abundant unique tag.
[d] [35].
[e] Tag frequency in hormone stimulated/starved samples.
[f] 2222 ESTs.
[g] 2236 ESTs.
[h] ~42,000 tags per library.
[i] ~62, 000 tags per library.
[j] nd, not done.

3.3. Comparison of EST and SAGE digital expression profiles

An alternate method of acquiring qualitative and quantitative transcript profiles is by the SAGE. Rather than producing gene tags of 300–500 nucleotides, the SAGE method generates sequence tags of approximately 10 nucleotides in length. This difference allows 10–30-fold more SAGE tags to be acquired per sequencing reaction, thus, deeper transcript profiles can be obtained more efficiently. However, the short tag length may introduce ambiguity when assigning a tag to a specific gene [21].

Data from two independent SAGE profiling experiments examining androgen-regulated gene expression in LNCaP cells were obtained from the SAGEmap website at NCBI.[14] Descriptions of the libraries indicated that one SAGE dataset, designated LNCaP(−)DHT/(+)DHT, was derived from LNCaP cells grown in hormone-depleted media for 3 months (LNCaP(−)DHT) and then stimulated with 1 nM DHT (LNCaP(+)DHT) for 24 h. Approximately 63,000 tags were sequenced from each library. The second SAGE dataset, LNCaP-T/-C, was derived from cells grown in hormone-depleted media for 5 days (LNCaP-C), then stimulated with 10^{-8} M R1881 for 24 h (LNCaP-T). Approximately, 42,000 tags were sequenced from each library. The distribution of expressed genes in each pair of SAGE libraries is given in Fig. 1B and C.

Theoretical and empirical data suggest that roughly 650,000 transcripts must be sampled to identify all but very rare mRNAs in the cell [22]. Thus, neither our study nor the SAGE datasets were large enough to thoroughly sample transcript diversity in the LNCaP cells, and neither dataset is capable of identifying differential gene expression among low abundance transcripts. Broadly, genes with a role in protein synthesis (ribosomal proteins and translation initiation factors) were the most abundant transcripts in both our EST data and the SAGE profiles. Interestingly, the EST approach identified approximately 200 transcript species

[14] http://www.ncbi.nlm.nih.gov/SAGE/.

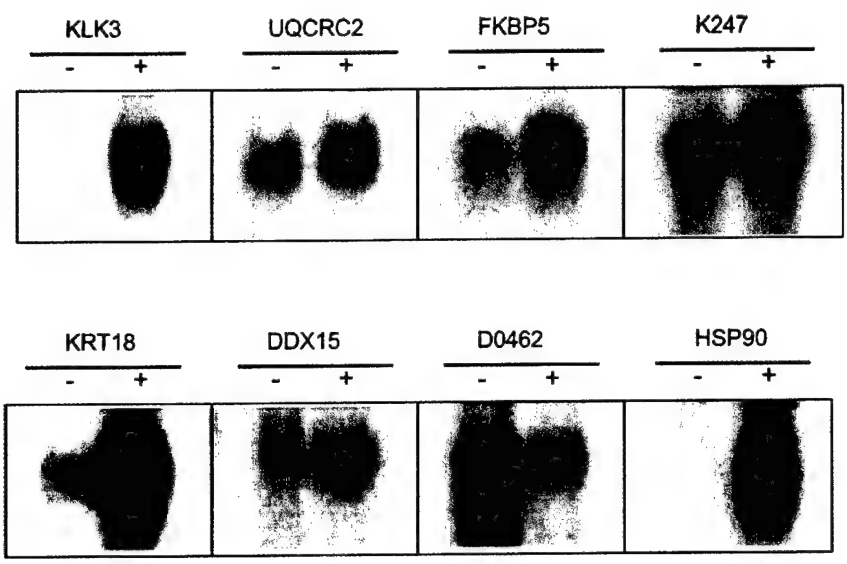


Fig. 4. Northern blots of eight androgen regulated genes predicted to be differentially expressed by virtual EST analysis. K247 is *DKFZP564K247* and D0462 is *DKFZP564D0462*. 'Minus', total RNA from androgen-starved LNCaP cells. 'Plus', total RNA from LNCaP cells treated with 1 nM R1881.

with corresponding Unigene entries that were not observed in the SAGE libraries. Conversely, the SAGE studies identified hundreds of transcripts that were not observed in the EST assemblies. Thus, these studies complement each other in creating an inventory representing the LNCaP cell transcriptome.

Transcripts with a high probability of differential expression between each pair of SAGE profiles were identified using the SAGEmap xProfiler. Despite a 10-fold difference in sample size, the SAGE and EST studies identified similar numbers of putative androgen-responsive genes (cut-off $P = 0.9$). In the EST analysis, 21 genes had a high probability

Table 3
Known androgen-response genes exhibiting differential expression in one or more libraries ($P \geq 0.6$)

Gene	ESTs			SAGE tag[a]	SAGE		
	No. of ESTs		Probability of differential expression[b]		Probability of differential expression[c,d]		Prostate-enriched[e]
	LNCaP01[f]	LNCaP02[g]			LNCaP-T/-C[h]	LNCaP(+)DHT/(−) DHT[i]	
CLTB	0	0	$0.00 < P < 0.10$	GGCTGGGCCT	$P = 0.45$ (3/0)	$P = 0.73$ (2112)	−
DBI	1	0	$0.50 < P < 0.60$	TGTTTATCCT	$P = 0.77$ (13/2)	$P = 0.03$ (20/18)	−
ENO1	6	3	$0.60 < P < 0.70$	GTGTCTCATC	$P = 0.13$ (9/12)	$P = 0.04$ (15/14)	−
KLK2	3	0	$0.80 < P < 0.90$	CTGTGGTTTA	$P = 0.39$ (2/0)	$P = 0.80$ (810)	+
	–	–	–	CTGTGGTTAA	–	$P = 0.76$ (14/3)	+
KLK3	29	1	$P > 0.99$	GGATGGGGAT	$P = 1.00$ (82/5)	$P = 0.25$ (63/36)	+
KLK4	2	0	$0.70 < P < 0.80$	AAATTGACCC	$P = 0.35$ (1/0)	$P = 0.51$ (2/8)	+
ODC1	4	1	$0.70 < P < 0.80$	TGCGTGGTCA	$P = 0.35$ (1/0)	–	−
	–	–	–	ATGCAGCCAT	–	$P = 0.11$ (7/7)	−
PDHA1	0	0	$0.00 < P < 0.10$	CAGTTTGTAC	$P = 0.60$ (5/0)	$P = 0.28$ (4/2)	−
PMEPA1[j]	1	0	$0.50 < P < 0.60$	TGATGTCTGG	$P = 1.00$ (29/1)	$P = 0.47$ (7/2)	+
TUBA1	4	1	$0.70 < P < 0.80$	GAGGAGGGTG	$P = 0.29$ (2/4)	$P = 0.44$ (5/13)	−
UGT2B17	0	0	$000 < P < 0.10$	GAGGGTTTTA	$P = 0.62$ (0/5)	$P = 0.40$ (4/1)	−
VEGF	1	1	$0.00 < P < 0.10$	TTTCCAATCT	$P = 0.29$ (1/2)	$P = 0.69$ (610)	−

[a] Most abundant unique tag.
[b] [11].
[c] [35].
[d] Tag frequency in hormone stimulated/starved cells.
[e] More abundant in the prostate than in most other tissues.
[f] 222 ESTs.
[g] 2236 ESTs.
[h] ~42,000 tags per library.
[i] ~62,000 tags per library.
[j] Tag inferred from [34].

of differential expression (9 up-regulated, 12 down-regulated) while 17 unique tags were identified in the SAGE LNCaP-T/-C study (6 up-regulated, 11 down-regulated), and 23 were identified in the SAGE LNCAP(+)DHT/(−)DHT study (17 up-regulated, 6 down-regulated). Surprisingly, with the exception of *KLK3/PSA*, all of the identified genes were different across the three datasets. *KLK3/PSA* had a high probability of differential expression in both our EST dataset ($P > 0.99$) and the LNCAP-T/-C dataset ($P = 1.0$). The only other potential androgen-regulated gene in the EST data that had a moderate probability of differential expression based on SAGE was FK506 binding protein 5 (*FKBP5*; $P = 0.66$, LNCaP-T/-C). The three genes that we confirmed to be differentially expressed by Northern blot analysis (keratin 18, 3-phosphoglycerate dehydrogenase, and *DKFZP564K247*) were not expressed at significantly different levels ($P < 0.30$) in the two SAGE datasets.

A review of published literature identified 75 genes reported to be androgen-responsive in one or more human tissues (see PEDB [15]). Twenty-three of these genes had corresponding EST tags; 47 had LNCaP-T/-C SAGE tags; and 55 had LNCaP(+)DHT/(−)DHT SAGE tags. Thus, SAGE sampling of 10-fold more transcripts only doubled the number of observed, previously-described, androgen-regulated genes. The genes identified in the EST dataset are not just a subset of those found in the larger SAGE datasets: *TMPRSS2*, a serine protease gene whose transcription is stimulated by androgen in LNCaP cells [23], was represented in the EST data, but not in the SAGE libraries. Only 12 of the 75 known androgen-response genes had even a moderate probability of differential expression ($P \geq 0.6$) in one or more datasets (Table 3), and there is no case where statistical predictions agree across all three data sets. Six of the twelve genes were predicted to be androgen inducible in the EST dataset, compared to five genes in the LNCaP-T/-C dataset and three in the LNCaP(+)DHT/(−)DHT dataset. The two SAGE studies, with similar numbers of tags, predicted completely different cohorts of up-regulated genes (Table 3).

4. Discussion

The identification and quantitation of the complement of genes expressed in a cell or tissue provides a framework for understanding biological properties and establishes a tool set for functional studies. Several methods have been developed for the comprehensive analysis of gene expression in complex biological systems. We have investigated the application of two procedures, EST profiling and SAGE, to characterize the transcriptome of prostate adenocarcinoma cells and to identify the cohort of genes regulated directly or indirectly by androgenic hormones. The EST profiles obtained from two LNCaP cDNA libraries identified 2486 distinct transcripts. Of these, 336 were expressed in common. The total number of transcripts, we identified in this study represents about 12–17% of the total complexity found in prostate epithelium [24] and likely includes all highly expressed, many moderately expressed and relatively few rarely expressed transcripts. Many of these genes were previously identified in other tissues, but were not known to be expressed in the prostate. In all, 252 new transcripts were identified that are not represented in any public database. Since over 2.2 million human ESTs are present in dbEST (release 081800), some of the unknown transcripts may be exclusively expressed in the prostate epithelium. These findings support the continued utility of cataloging transcripts from specialized tissue sources. These newly identified cDNAs can be tested for tissue-specific expression and can be used both to facilitate the identification of exons in the context of the human genome project and to enhance the positional cloning of prostate cancer susceptibility genes.

Androgens regulate numerous processes in prostate epithelial cells that include cell division, cell quiescence, apoptosis, lipid metabolism, and the production of specialized secretory proteins such as KLK3/PSA. Of the 2486 distinct transcripts identified in the LNCaP transcriptome, 364 (14%) showed at least a 2-fold difference in expression following exposure to androgens. Statistical analysis reduced this number to 21 genes with a high probability of differential expression ($P \geq 0.9$). Ten were further tested by Northern analysis which confirmed six were indeed transcriptionally regulated by androgen; *KLK3/PSA*, *FKBP5*, *KRT18*, *DDX15*, and *DKFZP564D0462*. In addition, *HSP90* was identified as an androgen-response gene by Northern blot analysis. These data identify five genes as new members of the androgen-response network, since only *KLK3/PSA* was previously known to be androgen-responsive. The lack of complete concordance between the digital expression results and Northern analysis can be partly explained by cross-hybridization to highly-homologous gene family members, alternative splicing events, and the lack of Northern sensitivity to alterations in low abundance transcripts.

The genes found in this study to be transcriptionally sensitive to androgen have diverse functions. *KLK3/PSA* is a highly abundant serine protease with known androgen-response elements in the promoter region [25] and prostate-enriched expression. Keratin 18 is a marker for prostate luminal cells [26] but is found in a variety of epithelia. The *DKFZP564D0462* gene encodes a putative seven transmembrane-domain protein that is expressed in a variety of tissues. The DEAD/H box polypeptide 15 gene is a putative RNA helicase similar to a yeast gene required for mRNA splicing [27]. Another RNA helicase, GRTH, is up-regulated in testis in response to androgen [19]. These genes may play a role in steroidogenesis or androgen-mediated stimulation of protein synthesis. *HSP90* binds and activates the androgen receptor. *FKBP5*, another gene predicted to be up-regulated in LNCaP cells, interacts with *HSP90* in func-

[15] http://www.pedb.org.

tionally mature progesterone complexes [28]. Hence, both *HSP90* and *FKBP5* may be up-regulated to facilitate signal transduction through the androgen receptor.

While general trends in gene expression were similar with respect to the overall effects of androgens, why was little concordance found between EST data and the SAGE data in terms of the expression of specific genes? In part this may be attributable to relatively small overall sample sizes and the limitations of statistical confidence. Cloning or sequencing biases could be unequally introduced by the experimental approaches, and ambiguity in SAGE tag assignment may affect a subset of genes. However, an alternative explanation is that each method accurately reflects the state of cellular gene expression, and the differences are attributable to the actual in vitro conditions. There will be some variation in transcript levels even under optimal conditions that may relate to cell density, growth media, and other factors. At present, we do not know the precise effects of protracted androgen starvation on LNCaP cells, but the extended starvation of cells used to create the LNCaP(+)DHT/(−)DHT libraries (3 months), could have selected for altered gene expression. In this regard, it is noteworthy that *KLK3/PSA*, one of the most abundant androgen regulated genes, was not differentially expressed in the LNCaP(+)DHT/(−)DHT dataset (Table 3). Cell-line history may also affect transcription. LNCaP may have undergone significant physiological adaptation and genomic change during maintenance in different laboratories. Esquenet et al. [29] observed a marked decrease in the ability of androgen to induce *KLK3/PSA* transcription in LNCaP cells of high passage number relative to cells of low passage number. And LNCaP cells can undergo "proliferative shut-off" in response to androgen [30]. These experimental differences may be analogous to the heterogeneity observed between individual cancers and may be reflected in the cellular transcriptomes assayed by digital-expression profiles.

Another intriguing possibility is that different androgens and androgen concentrations activate or repress subnetworks of the androgen-response program. Testosterone, DHT, and synthetic androgens such as R1881 induce a concentration-dependent biphasic growth response in LNCaP cells that may be influenced by the relative activities of growth-promoting and growth-suppressing genes [31]. Different ligands or ligand concentrations may recruit distinct AR co-activator molecules that dictate the subset of genes to be activated [32,33]. Of interest, a report describing the cloning and characterization of the gene corresponding to the SAGE tag exhibiting the greatest androgen-induction (29-fold) in the LNCaP(+)DHT/(−)DHT SAGE dataset was recently published [34]. By Northern analysis, the expression of this gene, *PMEPA1*, was shown to increase only 2-fold with 10^{-10} M R1881, but nearly 5-fold with 10^{-8} M R1881; the concentration used in the SAGE experiments. The 10^{-9} M R1881 concentration used in our EST experiments did not induce a detectable increase in *PMEPA1* EST frequency.

At present, financial and technological barriers make it impractical to simultaneously test all known genes for expression in the prostate. Inventories of genes from cell lines such as LNCaP, which are used extensively as model systems for studying prostate cancer, can help alleviate this problem by identifying the subset of genes of relevant to the biological system under study. Additional SAGE and EST data are needed to identify rare transcripts and to increase statistical power required for robust digital expression studies. In addition to their demonstrated utilities as gene discovery and analysis tools, the digital expression profiling methods used here can also greatly facilitate the construction of microarray-based reagents suitable for applications where higher throughput is required.

Acknowledgements

We thank Barbara Trask for critical review of the manuscript, David Haynor and Roger Bumgarner for helpful discussions, and Sue Heiner for manuscript preparation. This work is supported in part by grants (DAMD 17-98-1-8499 and PC991274) from the Department of Defense, a Grant (CA75173) from the National Cancer Institute, and the CaPCURE foundation.

References

[1] A.O. Brinkmann, L.J. Blok, P.E. de Ruiter, P. Doesburg, K. Steketee, C.A. Berrevoets, J. Trapman, Mechanisms of androgen receptor activation and function, J. Steroid Biochem. Mol. Biol. 69 (1999) 307–313.

[2] J. Trapman, K.B. Cleutjens, Androgen-regulated gene expression in prostate cancer, Seminars Cancer Biol. 8 (1997) 29–36.

[3] B. Ewing, P. Green, Analysis of expressed sequence tags indicates 35,000 human genes, Nat. Genet. 25 (2000) 232–234.

[4] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.

[5] V.E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M.A. Basrai, D.E. Bassett, P. Hieter, B. Vogeistein, K.W. Kinzler, Characterization of the yeast transcriptome, Cell 88 (1997) 243–251.

[6] G.G. Lennon, H. Lehrach, Hybridization analyses of arrayed cDNA libraries, Trends Genet. 7 (1991) 314–317.

[7] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.

[8] L. Wodicka, H. Dong, M. Mittmann, M.-H. Ho, D.J. Lockhart, Genome-wide expression monitoring of *Saccharomyces cerevisae*, Nature Biotechnol. 15 (1997) 1359–1367.

[9] M.D. Adams, A.R. Kerlavage, R.D. Fleischman, R.A. Fuldner, C.J. Bult, N.H. Lee, E.F. Kirkness, K.G. Weinstock, J.D. Gocayne, O. White, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence, Nature 377 (Suppl. 28) (1995) 3–174.

[10] V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, Serial analysis of gene expression, Science 270 (1995) 384–387.

[11] S. Audic, J.M. Claverie, The significance of digital gene expression profiles, Genome Res. 7 (1995) 986–995.

[12] V. Hawkins, D. Doll, R. Bumgarner, T. Smith, C. Abajian, L. Hood, P.S. Nelson, PEDB: the prostate expression database, Nucl. Acids Res. 27 (1999) 204–208.
[13] G.J. van Steenbrugge, M. Groen, J.W. van Dongen, J. Bolt, H. van der Korput, J. Trapman, M. Hasenson, J. Horoszewicz, The human prostatic carcinoma cell line LNCaP and its derivatives: an overview, Urol. Res. 17 (1959) 71–77.
[14] T. Maniatis, E.F. Fritsch, J. Sambrook, Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1982.
[15] U. Gubler, B.J. Hoffman, A simple and very efficient method for generating cDNA libraries, Gene 25 (1983) 263–269.
[16] X. Huang, An improved sequence assembly program, Genomics 33 (1996) 21–31.
[17] N.D. Hastie, J.O. Bishop, The expression of three abundance classes of messenger RNA in mouse tissue, Cell 9 (1976) 761–774.
[18] P.S. Nelson, W.L. Ng, M. Schummer, L.D. True, A.Y. Liu, R.E. Bumgarner, C. Ferguson, A. Dimak, L. Hood, An expressed-sequence-tag database of the human prostate: sequence analysis of 1168 cDNA clones, Genomics 47 (1998) 12–25.
[19] P.Z. Tang, C.H. Tsai-Morris, M.L. Dufau, A novel gonadotropin-regulated testicular RNA helicase: a new member of the dead-box family, J. Biol. Chem. 274 (1999) 37932–37940.
[20] K. Ozawa, Y. Murakami, T. Eki, E. Soeda, K. Yokoyama, Mapping of the gene family for human heat-shock protein 90 alpha to chromosomes 1,4,11, and 14, Genomics 12 (1992) 214–220.
[21] J. Stollberg, J. Urschitz, Z. Urban, C.D. Boyd, A quantitative evaluation of SAGE, Genome Res. 10 (2000) 1241–1248.
[22] V.E. Velculescu, S.L. Madden, L. Zhang, A.E. Lash, J. Yu, C. Rago, A. Lal, C.J. Wang, G.A. Beaudry, K.M. Ciriello, B.P. Cook, M.R. Dufault, A.T. Ferguson, Y. Gao, T.C. He, H. Hermeking, S.K. Hiraldo, P.M. Hwang, M.A. Lopez, H.F. Luderer, B. Mathews, J.M. Petroziello, K. Polyak, L. Zawel, K.W. Kinzler, Analysis of human transcriptomes, Nat. Genet. 23 (1999) 387–388.
[23] B. Lin, C. Ferguson, J.T. White, S. Wang, R. Vessella, L.D. True, L. Hood, P.S. Nelson, Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2, Cancer Res. 59 (1999) 4180–4184.
[24] L. Zhang, W. Zhou, V.E. Velculescu, S.E. Kern, R.H. Hruban, S.R. Hamilton, B. Vo-gelstein, K.W. Kinzler, Gene expression profiles in normal and cancer cells, Science 276 (1997) 1268–1272.
[25] K.B. Cleutjens, C.C. van Eekelen, H.A. van der Korput, A.O. Brinkmann, J. Trap-man, Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter, J. Biol. Chem. 271 (1996) 6379–6388.
[26] E.R. Sherwood, L.A. Berg, N.J. Mitchell, J.E. McNeal, J.M. Kozlowski, C. Lee, Differential cytokeratin expression in normal, hyperplastic and malignant epithelial cells from human prostate, J. Urol. 143 (1990) 167–171.
[27] O. Imamura, M. Sugawara, Y. Furuichi, Cloning and characterization of a putative human RNA helicase gene of the DEAH-box protein family, Biochem. Biophys. Res. Commun. 240 (1997) 335–340.
[28] S.C. Nair, R.A. Rimerman, E.J. Toran, S. Chen, V. Prapapanich, R.N. Butts, D.F. Smith, Molecular cloning of human *FKBP51* and comparisons of immunophilin interactions with *Hsp90* and progesterone receptor, Mol. Cell. Biol. 17 (1997) 594–603.
[29] M. Esquenet, J.V. Swinnen, W. Heyns, G. Verhoeven, LNCaP prostatic adenocarcinoma cells derived from low and high passage numbers display divergent responses not only to androgens but also to retinoids, J. Steroid Biochem. Mol. Biol. 62 (1997) 391–399.
[30] P. Geck, J. Szelei, J. Jimenez, T.M. Lin, C. Sonnenschein, A.M. Soto, Expression of novel genes linked to the androgen-induced, proliferative shutoff in prostate cancer cells, J. Steroid Biochem. Mol. Biol. 63 (1997) 211–218.
[31] E.G. Langeler, C.J. van Uffelen, M.A. Blankenstein, G.J. van Steenbrugge, E. Mulder, Effect of culture conditions on androgen sensitivity of the human prostatic cancer cell line LNCaP, Prostate 23 (1993) 213–223.
[32] P.W. Hsiao, T.H. Thin, L.D. Lin, C. Chang, Differential regulation of testosterone vs. Salpha-dihydrotestosterone by selective androgen response elements, Mol. Cell. Biochem. 206 (2000) 169–175.
[33] S. Yeh, H.C. Chang, H. Miyamoto, H. Takatera, M. Rahman, H.Y. Kang, T.H. Thin, H.K. Lin, C. Chang, Differential induction of the androgen receptor transcriptional activity by selective androgen receptor coactivators, Keio. J. Med. 48 (1999) 87–92.
[34] L.L. Xu, N. Shanmugam, T. Segawa, I.A. Sesterhenn, D.G. McLeod, J.W. Moul, S. Srivastava, A novel androgen-regulated gene, *PMEPA1*, located on chromosome 20q13 exhibits high level expression in prostate, Genomics 66 (2000) 257–263.
[35] A.E. Lash, C.M. Tolstoshev, L. Wagner, G.D. Schuler, R.L. Strausberg, G.J. Rig-gins, S.F. Altschul, SAGEmap: a public gene expression resource, Genome Res. 10 (2000) 1051–1060.

The Prostate 50:92–103 (2002)
DOI 10.1002/pros.10036

Human Prostate Epithelial Cell-Type cDNA Libraries and Prostate Expression Patterns

Alvin Y. Liu,[1,3]* Peter S. Nelson,[2] Ger van den Engh,[3] and Leroy Hood[3]

[1]Department of Urology, University of Washington, Seattle, Washington
[2]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington
[3]Institute for Systems Biology, Seattle, Washington

BACKGROUND. Transcriptome analysis is a powerful approach to uncovering genes responsible for diseases such as prostate cancer. Ideally, one would like to compare the transcriptomes of a cancer cell and its normal counterpart for differences.
METHODS. Prostate luminal and basal epithelial cell types were isolated and cell-type-specific cDNA libraries were constructed. Sequence analysis of cDNA clones generated 505 luminal cell genes and 560 basal cell genes. These sequences were deposited in a public database for expression analysis.
RESULTS. From these sequences, 119 unique luminal expressed sequence tags (ESTs) were extracted and assembled into a luminal-cell transcriptome set, while 154 basal ESTs were extracted and assembled into a basal-cell set. Interlibrary comparison was performed to determine representation of these sequences in cDNA libraries constructed from prostate tumors, PIN, cell lines.
CONCLUSIONS. Our analysis showed that a significant number of epithelial cell genes were not represented in the various transcriptomes of prostate tissues, suggesting that they might be underrepresented in libraries generated from tissue containing multiple cell types. Although both luminal and basal cell types are epithelial, their transcriptomes are more divergent from each other than expected, underscoring their functional difference (secretory vs. nonsecretory). Tumor tissues show different expression of luminal and basal genes, with perhaps a trend towards expression of basal genes in advanced diseases. *Prostate 50: 92–103, 2002.*

KEY WORDS: **prostate epithelial cell types; cell-type specific transcriptomes**

INTRODUCTION

The major constituent cell types of the adult prostate are the luminal epithelial, basal epithelial, and stromal fibromuscular cells [1]. Prostatic epithelial and stromal cells have different densities and can be separated by centrifugation in density gradients [2]. Because of their stem cell-like properties such as proliferative potential and differentiative plasticity, basal cells are postulated to be the likely progenitors of luminal cells [3]. Luminal cells are the terminally differentiated cells that perform the secretory function of the gland. Stromal fibromuscular cells have an important role in the induction of epithelial cell differentiation [1]. Synthesis of the abundant protein prostate-specific antigen (PSA) by luminal cells was shown to require the presence of stromal cells [4].

For unknown reasons, prostate epithelial cells are prone to malignant transformation. The advent of computational biology and genomics provides us with the means of analyzing and comparing repertoires of expressed genes or transcriptomes from

Abbreviation: EST, expressed sequence tag; PIN, prostatic intraepithelial neoplasia

Grant sponsor: CaP CURE Foundation.

*Correspondence to: Alvin Y. Liu, Ph.D., Department of Urology, Box 356510, University of Washington, Seattle, WA 98195.
E-mail: aliu@u.washington.edu

Received 8 May 2001; Accepted 9 October 2001

different cells. One approach is to first identify the genes associated with the cancer phenotype. This approach starts with the construction of representative cDNA libraries, followed by large-scale DNA sequencing of many cDNA clones and some type of comparison or subtractive analysis. Standard methods of cDNA library construction entail the use of tissue samples of several hundred milligrams. An inherent drawback in the use of tissue is heterogeneity, as the cell-type composition invariably differs from tissue to tissue (not always revealed by histomorphology). Thus, there is the likelihood that a difference in gene expression reflects different proportions of normal cell types rather than a true cancer difference. Laser-capture microdissection is a technical advance that permits a more precise excision of targeted tissue specimens [5] and many useful cDNA libraries have been constructed from specimens thus procured [6]. We have developed a complementary approach by employing flow cytometry to sort single-cell populations defined by their differentially expressed cluster designation (CD) antigens [4]. CD antigens are cell-surface molecules (http://www.ncbi.nlm.nih.gov/prow/). We examined prostatic expression of over 130 such CD antigens and nearly every cell type in the prostate can be identified by specific sets of CD antibodies. Cell populations sorted by CD expression can be used in the construction of cell-type-specific cDNA libraries. A comparison of the gene sequences cloned in these libraries should allow for the molecular characterization of the cellular phenotype and cell-type-specific transcriptomes of the two prostate epithelial cell types.

At present, DNA sequences of prostate cDNA are annotated in a prostate expression database (PEDB, http://www.pedb.org) assembled by us [7]. PEDB is a curated relational database containing over 40 prostate cDNA libraries identified by their tissue or cell source and 65,000 ESTs that are clustered into 21,000 species or genes. Tools to interrogate the expression of any sequence and its abundance among different libraries are built into the database.

MATERIALS AND METHODS

Cell-Type Analysis and Cell Isolation by Flow Cytometry

R-phycoerythrin (PE)-conjugated αCD44 and αCD57 monoclonal antibodies were obtained from PharMingen (San Diego, CA) and Sigma (St. Louis, MO), respectively. For flow analysis, prostate tissue specimens were minced and digested by collagenase in RPMI1640 media supplemented with 5% FBS and 10^{-8} M dihydrotestosterone at 37°C overnight. The cell suspension was then aspirated through a syringe and resuspended in 0.1% BSA-HBSS. Aliquots were labeled with either αCD57-PE or αCD44-PE. Positive cells were scored as events that registered outside the unstained and autofluorescent populations (visualized when no antibody or an irrelevant antibody was used). For flow sorting, prostate tissue specimens were digested by collagenase as above and loaded onto a Percoll discontinuous density gradient to separate the epithelial cells from the stromal fibromuscular cells. The epithelial fraction (containing both basal and luminal cells) was aspirated off the gradient and resuspended in 0.1% BSA-HBSS for labeling with either αCD57-PE for sorting of luminal cells or αCD44-PE for sorting of basal cells. To maximize yield, PE-conjugated antibodies were preferred over fluorescein isothiocyanate (FITC)-conjugated ones. Cells were collected in RPMI1640, pelleted, and lysed in STAT60 (Tel-Test "B," Friendswood, TX) for RNA isolation. A high-speed flow cytometer built in-house was used in these experiments; its features were described previously [4].

cDNA Library Construction

RNA from 200,000 to 400,000 sorted CD57- or CD44-positive cells was converted into cDNA by the SMART cDNA cloning technique (CLONTECH, Palo Alto, CA) as described previously [8]. The cDNA molecules were cloned into the bacterial vector pSPORT (Gibco-BRL, Bethesda, MD) and transformed into DH5α bacteria. Random bacterial colonies were chosen and recombinant clones were screened by PCR with DNA primers complementary to sequences flanking the cloning site. Clones with insert were sequenced and the resultant DNA sequences were deposited in PEDB and annotated. Sequence data manipulation is described in Ref. 7. The luminal-cell library was coded as UW PLC01 and the basal-cell library was coded as UW PBC01 in PEDB.

Interlibrary EST Analysis

A virtual expression analysis tool (VEAT) was incorporated into PEDB for interlibrary comparison and was used to analyze transcript abundance and differential expression. The size of the various libraries ranged from 100–6,000 sequences. For any pair of libraries selected for analysis a command to display common sequences between the two was executed. The visual output was a dot plot with each dot representing an EST. By clicking on the dot, the identity of the EST represented was retrieved and results of the comparisons were tabulated. Another sequence of commands under SEARCH was executed to determine the frequency of a particular EST among the different cDNA libraries.

TABLE I. Prostate Epithelial Cell-Type Transcriptome Sets, LC and BC

LC transcriptome-set

#204	2	109822 EST
#483*	3	calcium/calmodulin-dependent protein kinase (CaM kinase) IIγ
#663	1	131973 EST
#752	1	222399 EST weakly similar to *C. elegans* multiple EGF-like domain
#1042	1	KIAA0488 chromosome 1 transcript
#1470*	1	199638 EST†
#1824*	1	ribosomal protein L38
#2738	2	204335 EST
#2742*	2	108104 EST ubiquitin-conjugating enzyme E2L3
#2844	1	IL-1R-like†
#2972	1	H1 histone family member 2
#3004	1	204010 EST
#3079	1	H β58 homolog
#3289*	2	83006 EST moderately similar to *M. musculus* ganglioside-induced protein 3
#3454	1	63908 EST
#3522	1	T-cell activation protein EB1 family†
#3721	2	NADH dehydrogenase (ubiquinone) 1α subcomplex 6
#4040	2	91532 EST
#4383	2	*H. sapiens* clone 23675
#4473	1	chromosome 1 mRNA with similarities to BAT2
#4550*	3	endothelin receptor type A†
#4596	5	unassigned
#4656	2	86671 EST†
#4784	9	70732 EST
#4814	4	RING zinc finger (RZF)
#4978	1	mRNA of muscle specific gene M9
#5484	1	unassigned
#5550	3	clone 64K7 chromosome 20q11.21-11.23 translation initiation factor EIF2B2
#5578	1	181526 EST†
#5826	3	94722 EST†
#6132	3	171774 EST
#6145	1	124762 *H. sapiens* mRNA cDNA DKFZp566G163†
#6244	1	tip association protein
#6367	1	deleted in split-hand/split-foot 1 region
#6372	1	heat shock 105 kDα
#6395*	2	butyrate response factor 1 (EGF-response factor 1)
#6410	3	cell division cycle 27
#6662	3	110803 EST
#6866	3	186632 EST†
#6891	1	161489 EST
#6896	3	7535 EST highly similar to COBW-like placental protein
#7221*	6	H3 histone family 3B (H3.3B)
#7246	1	159392 EST
#7287	5	RAD21 *S. pombe* homolog
#7353	1	unassigned
#7790	1	translocation protein 1
#7797	1	small nuclear ribonucleoprotein D3
#7849	1	186632 EST†
#7882	3	heterochromatin protein HPIHs-γ
#7923	1	ATP synthase H⁺ transporting mitochondrial complex F0 subunit c isoform 1
#7978	1	proteasome (prosome macropain) subunit α type 2
#8531	1	thyroid receptor interacting protein 10 (CDC42-interacting)
#8813	1	KIAA0374 gene product†
#8825	2	nuclear protein marker for differentiated aortic smooth muscle†
#8859	1	hepatitis B virus x-interacting
#8907	1	glutathione requiring prostaglandin D synthase†
#8939*	2	proteoglycan 2 bone marrow (NK cell activator, eosinophil granule binding)†
#8946	1	12772 EST
#9038*	1	ATP synthase H⁺ transporting mitochondrial F0 complex subunit F6
#9065*	2	PTPRF interacting protein binding protein 2 (liprin β2)
#9074	1	cell division cycle 42 (GTP-binding)
#9091*	1	prothymosin α
#9106*	1	guanine nucleotide binding protein α inhibiting activity polypeptide 3
#9223*	2	ubiquitin-binding protein P62 phosphotyrosine independent ligand for Lck SH
#9269	1	STAT induced STAT inhibitor-4†
#9355	1	interferon-induced protein 17
#9407*	3	KIAA0266 gene product

TABLE I. (*Continued*)

#9645	1	163724 EST†
#9762*	1	tumor susceptibility gene 101
#10054	5	23044 EST†
#10164*	2	unassigned
#10231	2	human homolog of yeast mitochondrial copper recruitment gene
#10364*	1	153703 EST moderately similar to succinate dehydrogenase†
#10496	2	nuclear mitotic apparatus protein 1
#10505	1	LIM domain kinase 2†
#10561	1	unassigned†
#10615	1	208954 EST†
#10620	4	193898 EST†
#10734*	10	208189 mRNA cDNA DKFZp566O053†
#10777	1	general transcription factor IIH polypeptide 1†
#10824*	19	mitochondrial genome
#10932	3	146247 EST†
#10982	2	cDNA DKFZp564H2416
#11219	3	104215 EST
#11254*	2	ribosomal protein S6
#11507	1	basic transcription factor 3
#11510	1	MAX binding†
#11882	2	DR1-associated (negative cofactor 2α)
#12001	1	44163 EST highly similar to 13 kD differentiation-associated
#12059	2	placental growth factor vascular endothelial growth factor-related†
#12150	9	myosin light polypeptide regulatory non-sarcomeric
#12182	8	180145 EST†
#12440	2	25341 EST†
#12670	1	44017 EST
#12879*	1	SRB7 suppressor of RNA polymerase B yeast homolog
#12974	3	BH-protocadherin†
#13030	1	34060 EST†
#13144*	1	sin 3-associated
#13225	2	97058 EST highly similar to CMP-N-acetylneuraminic acid hydroxylase†
#13247	2	3385 EST
#13344	1	22964 EST
#13386	3	homolog of *S. cerevisiae* ufd2†
#13531	1	11411 EST
#13677	1	ferritin light
#14020	1	ATP synthase H^+ transporting mitochondrial F0 complex subunit c isoform 3
#14021	2	acetyl-Coenzyme A acetyltransferase 2 (CoA thiolase)
#14332	1	193330 EST
#14723*	1	59698 EST
#14756	2	glyoxalase 1
#14877	3	eukaryotic translocation initiation factor 1A
#14978	1	calmodulin 2 (phosphorylase kinase δ)
#15191	3	ribosomal protein L6
#15286	1	24156 EST weakly similar to transporter protein
#15415	3	SC35-interacting protein 1
#16091*	1	vimentin
#16185	1	transglutaminase 4
#16468	2	catenin α1
#16783	1	Williams-Beuren syndrome chromosome region 10
#16950	1	death-associated protein 6

TABLE I. (*Continued*)

BC transcriptome-set		
#138	1	65648 EST
#205	2	leucine rich repeat (in FLII) interacting protein 1
#483*	2	calcium/calmodulin-dependent protein kinase (CaM kinase) IIγ
#625	1	TAR (HIV) RNA-binding protein 1
#653	1	132055 EST
#724	1	KIAA0564 gene product
#822	2	POP4 (processing of precursor *S. cerevisiae*) homolog†
#835	3	ADP-ribosylation factor 1
#878	3	7862 EST weakly similar to *R. norvegicus* proline rich protein
#1278	1	197990 EST†
#1322	2	cDNA DKFZp564O0823
#1421	1	22209 EST†
#1470*	3	199638 EST†
#1557	4	eukaryotic translation initiation factor 3 subunit 6
#1567	1	nucleolin
#1735	3	thymosin β4 X chromosome
#1808	4	TGFβ receptor III (betaglycan)
#1824*	7	ribosomal protein L38
#1932	1	protein tyrosine phosphatase receptor type K
#1990	1	methionine aminopeptidase eIF-2-associated p67
#2123	3	ubiquitin C
#2742*	1	108104 EST ubiquitin-conjugating enzyme E2L3
#2804	1	ATPase Ca^{+2} transporting plasma membrane 1
#2970	2	ribosomal protein S10
#3065	5	50252 EST†
#3098	1	golgi autoantigen golgin subfamily b macrogolgin 1
#3286	2	cytochrome c oxidase subunit VIIb
#3289*	1	83006 EST moderately similar to *M. musculus* ganglioside-induced protein 3
#3410	2	mitochondrial enoyl Coenzyme A hydratase short chain 1
#3445	1	ribosomal protein L19
#3568	5	ubiquitin-conjugating enzyme E2 variant 1
#3575	1	hemopoietic progenitor homeobox HPX42B
#3654	1	neuroblastoma RAS viral oncogene homolog
#3681	9	novel centrosomal protein RanBPM
#3760	4	KIAA0666 gene product†
#3889	1	8454 EST highly similar to camp-dependent protein kinase type II-α regulatory
#3937	1	84359 mRNA for hypothetical protein
#3955	1	unassigned†
#4017	3	mRNA and cDNA clone EUROIMAGE 45620
#4119	1	CD63 antigen (melanoma antigen)
#4124	1	eukaryotic translation initiation factor 3 subunit 5 (ε)
#4200	1	A9A2BR11 (CAC)n/(GTG)n repeat-containing mRNA†
#4360	3	66295 EST weakly homologous of Drosophila discs large protein isoform†
#4416	1	splicing factor (CC1.3)
#4550*	1	endothelin receptor type A†
#4890	5	ribosomal protein S25
#4913	1	Kin17†
#4941	11	tumor rejection antigen (gp96) 1
#5017	1	KIAA0341 gene product†
#5143	1	IL-15Rα†
#5177	1	195568 EST highly similar to NF90 protein
#5493	1	191367 EST highly similar to *M. musculus* Dhm1 protein†
#5504	3	restin (Reed-Steinberg cell-expressed intermediate filament-associated)
#5528	1	3742 EST highly similar to *R. norvegicus* protein transport protein SEC61α
#5827	1	superoxide dismutase 1 soluble (amyotrophic lateral sclerosis 1)
#6121	4	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein θ
#6164	1	preprotein translocase
#6209	1	74375 EST
#6258	1	206950 EST
#6395*	1	butyrate response factor 1 (EGF-response factor 1)
#6468	1	clone 1183I21 on chromosome 20q1.2
#6552	1	*t*-complex-associated-testis-expressed 1
#6555	1	citrate synthase
#6576	1	PBX/knotted 1 homeobox 1†
#6689	12	ribosomal protein L5
#6861	1	ubiquitin-conjugating enzyme E2N (homologous to yeast UBC13)
#6914	4	186632 EST

TABLE I. (*Continued*)

#7020	1	57672 EST weakly similar to *M. musculus* FLI-LRR associated protein-1†
#7181	2	144183 EST†
#7221*	3	H3 histone family 3B (H3.3B)
#7401	1	proteasome (prosome macropain 26S subunit non-ATPase 7 (Mov34 homolog)
#7832	1	KIAA0741 gene product
#7970	1	ribosomal protein L35
#8011	3	ribosomal protein L32
#8101	1	aldolase B fructose-bisphosphate
#8199	1	cytochrome c oxidase subunit VIIa polypeptide 2 (liver)
#8268	1	186632 EST†
#8286	2	36475 EST
#8528	1	ribosomal protein L7a
#8567	1	DNA segment on chromosome X 648 expressed sequence
#8669	1	immunoglobulin λ gene cluster
#8760	2	heart mRNA for HSP90
#8844	5	146565 EST
#8939*	1	proteoglycan 2 bone marrow (NK cell activator, eosinophil granule binding)†
#8993	1	Janus kinase 1
#9038*	1	ATP synthase H^+ transporting mitochondrial F0 complex subunit F6
#9065*	1	PTPRF interacting protein binding protein 2 (liprin β2)
#9091*	1	prothymosin α
#9106*	1	guanine nucleotide binding protein α inhibiting activity polypeptide 3
#9131	2	lactate dehydrogenase B
#9223*	1	ubiquitin-binding protein P62 phosphotyrosine independent ligand for Lck SH
#9357	1	DEAD/H (asp-glu-ala-asp/his) box polypeptide 16
#9364	1	222903 EST
#9391	1	unassigned
#9407*	2	KIAA0266 gene product
#9408	3	mRNA for 23 kD highly basic protein
#9666	1	63288 EST
#9689	1	structural maintenance of chromosome (SMC) family member protein E†
#9762*	12	tumor susceptibility gene 101
#9781	7	ribosomal protein L44
#9786	2	prefoldin 1
#9815	1	KIAA0853 gene product
#9918	4	132785 EST weakly similar to *C. elegans* predicted protein F17C8.5
#10004	1	186632 EST†
#10051	1	23044 EST†
#10073	1	186632 EST†
#10074	3	186632 EST†
#10140	1	153197 EST†
#10319	1	E74-like factor 1 (ets domain transcription factor)
#10364*	1	153703 EST moderately similar to succinate dehydrogenase†
#10734*	11	208189 mRNA cDNA DKFZp566O053†
#10824*	6	mitochondrial genome
#10893	1	103657 EST weakly similar to CH-TOG protein†
#10928	2	116567 EST†
#10984	1	103493 EST†
#10987	2	23120 EST
#11023	1	115880 EST†
#11254*	10	ribosomal protein S6
#11484	2	101150 EST
#11742	2	9061 EST
#11907	1	103845 EST
#11967	1	ribosomal protein L23
#12117	2	high density lipoprotein binding
#12157	1	BC-2 protein mRNA
#12207	1	20100 EST†
#12291	1	DNA-directed polymerase β†
#12305	1	177181 EST†
#12542	1	11473 EST†
#12609	1	guanylate kinase 1
#12879*	1	SRB7 suppressor of RNA polymerase B yeast homolog
#13144*	1	sin 3-associated
#13153	1	small inducible cytokine A2 (monocyte chemotactic protein 1)
#13450	1	IL-8
#13674	5	H2A histone member P†
#13827	1	serine/threonine kinase 9
#14042	1	ribosomal protein L4

TABLE I. (*Continued*)

#14149	1	ribulose-5-phosphate-3-epimerase†
#14392	2	clone 414D7 on chromosome 22q13.2-13.33 homologous to *C. elegans* T21D12.4†
#14636	1	5243 EST moderately similar to *R. norvegicus* pIL2 hypothetical protein
#14723*	1	59698 EST
#14993	1	maternal G10 transcript
#15168	3	caspase 6 apoptosis-related cysteine protease
#15236	1	serine/threonine kinase 2
#15509	1	chemoattractant receptor-homologous molecule expressed on TH2 cells
#15561	2	59038 EST
#15922	1	224318 EST†
#16076	1	126075 EST weakly similar to *C. elegans* C33G8.2†
#16091*	1	vimentin
#16164	1	118036 EST†
#16288	1	194449 EST†
#16597	3	PRKC apoptosis WT1 regulator
#16655	1	186643 EST
#16778	5	173518 EST weakly similar to M-phase phosphoprotein 4
#16794	1	13015 EST highly similar to *M. musculus* DNA J protein homology MTJ1

A cluster ID number is assigned to each entry as listed in the first column. ID numbers marked by an asterisk are the 24 sequences common to both sets. The frequency (3,2, etc.) of each sequence in the library is indicated to the right. The gene identity of each sequence is in the third column, with entries marked by a dagger to denote those that are not found in the prostate cDNA libraries listed in Table III.

RESULTS

Luminal and Basal Cell-Type Transcriptomes

The two major epithelial cell types in the adult prostate were sortable into either the $CD57^+$ or $CD44^+$ populations. Virtually all noncancerous tissue specimens (unlike those of cancer tissue) examined contained both $CD57^+$ and $CD44^+$ cell types. The cDNA libraries made from sorted cells were designated as PLC01 for $CD57^+$ luminal cells and PBC01 for $CD44^+$ basal cells. Five hundred and five PLC and 560 PBC sequences were analyzed, from which 119 and 154 single ESTs were assembled, respectively. These gene sequences were collected as transcriptome-sets LC (luminal) and BC (basal). In the LC group, 55 sequences (46.2%) were represented in the library at a frequency of ≥ 2 and were scored as "abundant" species. The remaining 64 sequences (53.8%) with a frequency of 1 were scored as "rare" species. In the BC group, 55 sequences (35.7%) were scored as "abundant" species and 99 sequences (64.3%) were scored as "rare" species. Each gene sequence was assigned a cluster identity number (#1, #2, etc.). Table I lists these genes in order of their cluster numbers, along with their identity. The 24 genes common to both sets are highlighted by asterisks and, of these, 17 were matched to known genes and 7 to ESTs. Of the 95 genes in LC and 130 genes in BC the abundant species have a high potential of being cell-type-specific (e.g., #4784, #7287, #10054, #12150, #12182 in LC; #3065, #3568, #3681, #4941, #8844, #16778 in BC with frequencies greater than 4, Table II). The key point is that unique genes of the abundant species were distinctly different in the luminal and basal libraries, consistent with quite different patterns of gene expression, even with the small sample size. This suggested that many distinct clones were represented in the libraries. With a larger sampling size, many of the ESTs will still probably be

TABLE II. High, Abundance Transcripts

LC	BC
Unassigned #4596	Translation initiation factor 3
70732 EST	TGF receptor
RING zinc finger	50252 EST
RAD21 *S. pombe* homolog	Ubiquitin-conjugating enzyme variant
23044 EST	Novel centrosomal protein
193898 EST	KIAA0666
DKFZp566O053	Tumor rejection antigen (gp96)
Myosin light polypeptide regulatory	Tyrosine 3-monooxygenase
180145 EST	186632 EST
	146565 EST
	Tumor susceptibility gene 101
	132785 EST
	DKFZp566O053
	173518 EST

Listed are sequences that have a frequency ≥ 4 in these cDNA libraries (mitochondrial, ribosomal protein, histone sequences are not included). One, DKFZp566O053, is found in both transcriptome sets.

TABLE III. Prostate cDNA Libraries

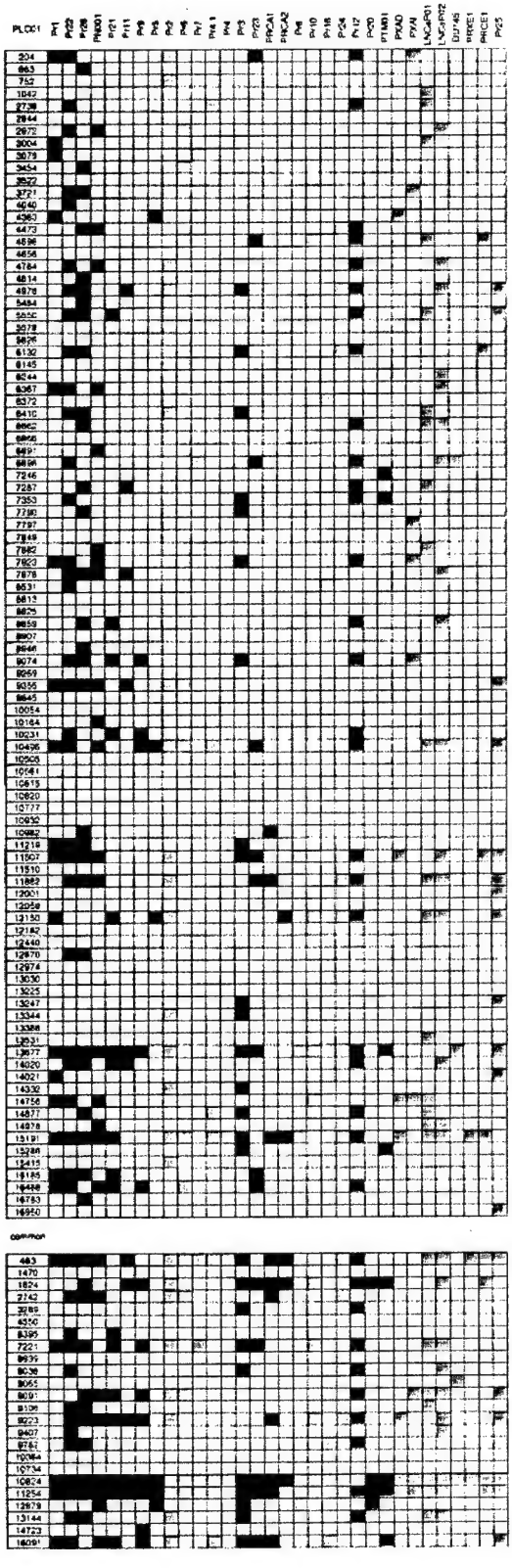
cDNA library	Sequences	Contigs	LC	BC
NCI CGAP Pr1 microdissected normal epithelium	5569	1916	18.5%	24%
NCI CGAP Pr22 normalized normal whole prostate	5767	3232	36.1%	40.3%
NCI CGAP Pr28 bulk subtracted normal prostate	4162	3188	33.6%	35.1%
UW PN001 normal whole prostate	2597	1349	21.9%	24%
NCI CGAP Pr21 non-normalized normal whole prostate	1237	691	15.1%	20.8%
NCI CGAP Pr11 microdissected normal epithelium	1334	669	10.1%	13%
NCI CGAP Pr9 microdissected normal epithelium	1057	606	11%	20.1%
NCI CGAP Pr5 microdissected normal epithelium	769	410	6.7%	10.4%
NCI CGAP Pr2 microdissected low-grade PIN	5529	2096	21%	30.5%
NCI CGAP Pr6 microdissected low-grade PIN	1436	765	9.2%	16.9%
NCI CGAP Pr7 microdissected low-grade PIN	459	265	5%	8.4%
NCI CGAP Pr4.1 microdissected high-grade PIN	1238	640	6.7%	17.5%
NCI CGAP Pr4 microdissected high-grade PIN	636	351	2.5%	10.4%
NCI CGAP Pr3 microdissected primary carcinoma	5057	1792	21.9%	29.2%
NCI CGAP Pr23 pooled broad spectrum primary carcinoma	987	606	9.2%	16.9%
UW PRCA1 primary carcinoma	666	383	10.9%	13%
UW PRCA2 primary carcinoma	369	194	4.2%	3.3%
NCI CGAP Pr8 microdissected primary carcinoma, invasive	1071	570	5%	14.3%
NCI CGAP Pr10 microdissected primary carcinoma, invasive	1120	540	8.4%	15.6%
NCI CGAP Pr16 microdissected primary carcinoma, invasive	539	231	4.2%	7.1%
NCI CGAP Pr24 HPV immortalized cell line from primary carcinoma, invasive	968	612	10.9%	13.6%
NCI CGAP Pr12 microdissected bone metastasis	4189	1778	29.4%	28.6%
NCI CGAP Pr20 microdissected liver metastasis	162	85	3.4%	3.9%
UW PTM01 liver metastasis	490	291	6.7%	11%
UW PXAD androgen dependent xenograft of primary carcinoma	605	368	5%	13%
UW PXAI androgen independent xenograft of primary carcinoma	449	297	7.6%	9.7%
UW LNCaP01 androgen stimulated LNCaP cells	2111	1114	20.2%	21.4%
UW LNCaP02 androgen starved LNCaP cells	2047	990	21.9%	24%
UW DU145 DU145 cancer cell line	237	143	3.4%	3.9%
UW PRXE1 SCID xenograft	309	43	2.5%	2.6%
UW PRCE1 cultured epithelium	596	280	6.7%	12.3%
NCI CGAP Pr25 HPV immortalized normal epithelial cell line	1408	753	15.1%	17.5%

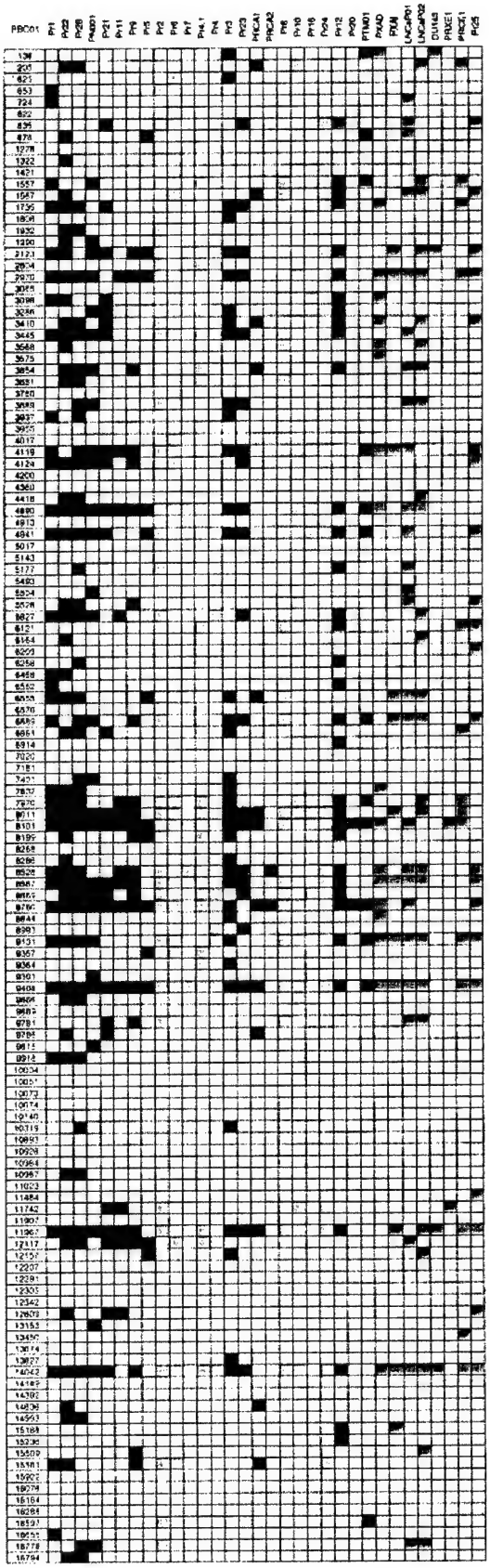
The cDNA libraries used in this report are grouped into NORMAL, PIN, CANCER, and CULTURED CELLS, The number of sequences deposited and genes in these libraries are given in the second and third columns, respectively. The percentages of sequence match between LC or BC and the other prostate cDNA libraries are listed in the last two columns.

uniquely expressed in each. At the time of writing, five EST sequences (#4596, #5484, #7353, #10164, #10561) in the luminal-cell set were unassigned by a Unigene annotation, while two (#3955, #9391) in the basal-cell set were unassigned. Among the others were ribosomal protein genes S6, S10, S25, L4, L5, L7a, L19, L23, L32, L35, L38, L44 in BC; S6, L6, L38 in LC; and one mitochondrial, three histone (H1.2, H2A.P, H3.3B) genes.

Representation of LC and BC Sequences in Prostate Libraries

The LC and BC transcriptome-sets were compared to gene sequences of various cDNA libraries available in PEDB. The libraries and tissue sources from which they were made are identified in Table III. The 32 libraries selected were grouped into four cohorts of 1) normal prostate; 2) prostate intraepithelial neoplasia (PIN); 3) prostate carcinoma, cancer cell lines and xenografts; and 4) cultured epithelial cells. Results of the interlibrary comparisons are graphically presented in Figure 1. Not represented in any of the other library sets (blank boxes in Fig. 1) were 33 or 27.7% (33/119) LC genes, which included genes encoding IL-1R-like protein, T-cell activation EB1, prostaglandin synthase, STAT inhibitor, LIM domain kinase, transcription factor, MAX binding protein, placental growth factor, protocadherin, endothelin receptor A, proteoglycan 2; and 42 or 27.3% (42/154) BC genes, which included ones encoding Kin 17, IL-15Rα, knotted 1, SMC


PLCG1
common


PBCG1

protein, DNA polymerase, histone, ribulose-5-phosphate-3-epimerase, endothelin receptor A, proteoglycan 2. A majority had an abundance frequency of 1 except EST 208189 (#10734) (Table I). The number of genes in these libraries ranged from 43 (PRXE1) to 3,232 (Pr22) species (see Table III).

When the LC and BC transcriptome-sets were matched against the other libraries in the prostate database, the percentage of matches, as expected, increased with the size of the library chosen, as tabulated in Table III. The match percentages ranged from 36.1% LC and 40.3% BC (including "rare" as well as "abundant" species) in Pr22 with 3,232 genes to 6.7% LC and 10.4% BC in Pr5 with 410 genes. These matches were done to characterize the cell types, luminal- or basal-like, that populate the diseased tissues as compared to normal tissue, which has both cell types.

For libraries of low-grade (Pr2, Pr6, Pr7) and high-grade (Pr4.1, Pr4) PIN (histologically discernible abnormalities that are considered to be precancerous), the average percentage difference between the higher BC and lower LC representation was 7.8% (6.9% for low-grade and 9.3% for high-grade), almost twofold as much as the value observed for normal prostate.

For libraries of carcinoma, the average difference between the BC and LC match percentages was 4.1% in *primary carcinoma* libraries and 6.5% in *primary carcinoma invasive* libraries. The difference was 2.7% for the library of a cell line derived from primary carcinoma invasive. Unlike most other comparisons, there was about equal representation of LC and BC sequences in the bone metastasis library Pr12. This ratio was also noted for libraries of prostate cancer cell lines and xenografts except PXAD. There was a higher BC representation for libraries of cultured cells.

Not found in the PIN and cancer libraries were the following LC sequences, with their abundance frequency in parentheses: #663 (1), #4040 (2), #4814 (4), #5484 (1), #6891 (1), #8946 (1), #10164 (2), #12670 (1), #16783 (1), #6395 (2), #14723 (1); and BC sequences: #1322 (2), #6468 (1), #9391 (1), #9815 (1), #9918 (4), #10987 (2), #13153 (1), #12609 (1), #14993 (1), #16655 (1), #16794 (1). Five LC sequences in *primary carcinoma invasive* [#4473 (1), #6372 (1), #8531 (1), #9762 (1), #14756 (2)] were not represented in the larger pool of sequences of *primary carcinoma*. And 11 [#1557 (4), #3098 (1), #3568 (5), #3681 (9), #5528 (1), #6121 (4), #9666 (1), #9762 (12), #9781 (7), #11484 (2), #11742 (2)] BC sequences in *primary carcinoma invasive* were not represented in *primary carcinoma*. Note the increase in genes of higher abundance. Three in the latter group (#3568, #9666, and #11484) showed an increased representation in libraries derived from tissues diagnosed as advanced diseases. One (#6121) was found in the library of a small-cell cancer xenograft (UW PRCA3).

DISCUSSION

Prostate cell-type transcriptomes represent important databases by which to study differential gene expression of cell lineages in development and cancer. In development, luminal cells are thought to differentiate from basal cells. By comparing the transcriptomes of these two cell types we can identify genes that are differentially expressed between them. These genes can be used as probes to study the neoplastic process since cancer is in some aspect a result of derangement in the cellular differentiation process.

For cDNA library construction, the two epithelial populations were isolated by their differentially expressed cell surface molecules, CD44 and CD57. There is some confusion in the literature regarding the cell-type specificity of the CD44 antigen. Based solely on immunohistochemistry, some investigators reported that both basal and luminal cells were positive for CD44 [9,10]. We and others [11,12] have shown that CD44 expression was localized to the basal cells. The discordance could perhaps be attributed to the antibody clones and immunostaining conditions used. We have also used cell sorting and RT-PCR to demonstrate the absence of CD44 mRNA expression in $CD57^+$ luminal cells [4].

Few experimental analyses have been carried out to determine the degree of difference between the transcriptomes of basal and luminal cells. A comparative analysis of cell-type-specific surface molecules showed that only a third of the epithelial-positive molecules were shared between the two cell types [13]. It is also quite clear that the two cell types are functionally different. If 25% is the estimated difference between the transcriptomes of fibroblasts and lymphocytes, ~2% that between those of T and B lymphocytes [14], then that for luminal and basal cells may lie between these two values. If it is 10% then 10–15 genes in the transcriptome-sets are probably cell-type-specific. If we assume that differentially expressed genes are more likely to be in the moderate

Fig. 1. LC and BC representation in prostate cDNA libraries. The LC and BC genes are placed by their cluster ID number. The various cDNA libraries are identified on the top of the grid pattern. Presence in a particular library is indicated by colored boxes: black for *normal*, rose for *PIN*, red for *primary carcinoma*, light orange for *primary carcinoma invasive*, blue for *metastasis*, lavender for *xenografts and cancer cell lines*, and lime for *cultured cells*.

and high abundance classes, then the likelihood of their being preferentially cloned in the libraries is increased. Hence, although our transcriptome sets are small the interlibrary comparisons using them would yield meaningful results.

In cancer, cell-type-specific ESTs can be used to examine gene expression of primary tumors and metastases. From our cancer cell-type analysis of tumor specimens we found that, whereas most primary tumors contained $CD57^+$ cancer cells, several metastases analyzed by us contained primarily $CD44^+$ cancer cells [15]. An association between CD44 expression and the invasive phenotype can also be made out from database analysis. The frequency of CD44 EST in the *primary carcinoma invasive* library Pr8 is 0.18, compared to 0.04 in the *primary carcinoma* library Pr3. The value of 0.18 is comparable to that of 0.16 in the *cultured epithelial cells* library PRCE1. We have shown by immunocytochemistry that nearly every cell in culture is positive for CD44 expression [16]. It is therefore possible that this particular primary carcinoma, characterized as invasive, contained a high proportion of CD44-positive cancer cells and presumably a higher BC representation, as indicated by our analysis. As with the two normal epithelial cell types, cancer cell types can be isolated by flow cytometry from the appropriate tumor sources for cDNA libraries and transcriptomes.

The presumed premalignant abnormality, PIN, appears to have a higher representation of BC than LC sequences from our analysis. The bias is more pronounced for high-grade PIN, which has a strong association with cancer [17]. A higher BC representation would suggest that PIN lesions are populated by "basal cell-like" cells. The presence in PIN of basal cell markers such as the RNA component of telomerase hTR [18], interleukin-6 [19], and bcl-2 [20] lends support to this suggestion. The use of BC and LC gene probes, along with CD antibodies, to determine the cell type composition of PIN lesions will clarify the lineage relationship of PIN cells.

In conclusion, we think that cell-type-specific cDNA libraries are vital to understanding the genetic mechanism of prostate cancer development. A normal prostate library made from tissue samples contains sequences from at least four cell types—luminal epithelial, basal epithelial, stromal, and white blood cells ($CD45^+$ or $CD43^+$, Ref. 13). Thus, from a library of 4,000 sequences only 1,000 may represent the transcriptome of, say, luminal cells. Consequently, it is not surprising that a significant number of LC or BC sequences are not found in the database. A prostate cancer library, on the other hand, contains sequences from at least three cell types—cancer epithelial, stromal, and white blood cells. Comparative analysis between these "tissue" libraries would likely yield many false-positives. With CD cell surface markers identified for most, if not all, prostate normal and diseased cell types [13], cDNA libraries can be constructed for any relevant cell type that can be sorted by flow.

ACKNOWLEDGMENTS

We thank Dr. Kristen Brubaker for comments on the manuscript.

REFERENCES

1. Cunha GR, Alarid ET, Turner T, Donjacour AA, Boutin EL, Foster BA. Normal and abnormal development of the male urogenital tract. Role of androgens, mesenchymal-epithelial interactions, and growth factors. J Androl 1992;13:465–475.
2. Kassen A, Sutkowski DM, Ahn H, Sensibar JA, Kozlowski JM, Lee C. Stromal cells of the human prostate: initial isolation and characterization. Prostate 1996;28:89–97.
3. Bonkhoff H, Stein U, Remberger K. Multidirectional differentiation in the normal, hyperplastic, and neoplastic human prostate: simultaneous demonstration of cell-specific epithelial markers. Hum Pathol 1994;25:42–46.
4. Liu AY, True LD, LaTray L, Nelson PS, Ellis WJ, Vessella RL, Lange PH, Hood L, van den Engh G. Cell-cell interaction in prostate gene regulation and cytodifferentiation. Proc Natl Acad Sci USA 1997;94:10705–10710.
5. Fend F, Emmert-Buck MR, Chuaqui R, Cole K, Lee J, Liotta LA, Raffeld M. Immuno-LCM: laser capture microdissection of immunostained frozen sections for mRNA analysis. Am J Pathol 1999;154:61–66.
6. Best CJ, Gillespie JW, Englert CR, Swalwell JI, Pfeifer J, Krizman DB, Petricoin EF, Liotta LA, Emmert-Buck MR. New approaches to molecular profiling of tissue samples. Anal Cell Pathol 2000;20:1–6.
7. Hawkins V, Doll D, Bumgarner R, Smith T, Abajian C, Hood L, Nelson PS. PEDB: the prostate expression database. Nucl Acids Res 1999;27:204–208.
8. Nelson PS. Single-cell cDNA libraries. In: Innis MA, editor. PCR applications. New York: Academic Press; 1999. p 307–328.
9. De Marzo AM, Bradshaw C, Sauvageot J, Epstein JI, Miller GJ. CD44 and CD44v6 downregulation in clinical prostatic carcinoma: relation to Gleason grade and cytoarchitecture. Prostate 1998;34:162–168.
10. Noordzij MA, van Steenbrugge GJ, Schröder FH, van der Kwast TH. Decreased expression of CD44 in metastatic prostate cancer. Int J Cancer 1999;84:478–483.
11. Paradis V, Eschwège P, Loric S, Dumas F, Ba N, Benoît G, Jardin A, Bedossa P. De novo expression of CD44 in prostate carcinoma is correlated with systemic dissemination of prostate cancer. J Clin Pathol 1998;51:798–802.
12. Fry PM, Hudson DL, O'Hare MJ, Masters JRW. Comparison of marker protein expression in benign prostatic hyperplasia *in vivo* and *in vitro*. BJU Int 2000;85:504–513.
13. Liu AY, True LD. Characterization of prostate cell types by CD cell surface molecules. Am J Pathol (in press).
14. Hedrick SM, Cohen DI, Nielsen EA, Davis MM. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature 1984;308:149–153.

15. Liu AY, True LD, LaTray L, Ellis WJ, Vessella RL, Lange PH, Higano CS, Hood L, van den Engh G. Analysis and sorting of prostate cancer cell types by flow cytometry. Prostate 1999;40: 192–199.
16. Liu AY, Peehl DM. Characterization of cultured human prostatic epithelial cells by cluster designation antigen expression. Cell Tissue Res 2001;305:389–397.
17. Bostwick DG. Prostatic intraepithelial neoplasia is a risk factor for cancer. Semin Urol Oncol 1999;17:187–198.
18. Paradis V, Dargère D, Laurendeau I, Benoît G, Vidaud M, Jardin A, Bedossa P. Expression of the RNA component of human telomerase (hTR) in prostate cancer, prostatic intraepithelial neoplasia, and normal prostate tissue. J Pathol 1999; 189:213–218.
19. Hobisch A, Rogatsch H, Hittmair A, Fuchs D, Bartsch G, Klocker H, Bartsch G, Culig Z. Immunohistochemical localization of interleukin-6 and its receptor in benign, premalignant and malignant prostate tissue. J Pathol 2000;191: 239–244.
20. Baltaci S, Orhan D, Özer G, Tolunay Ö, Göüs O. Bcl-2 protooncogene expression in low- and high-grade prostatic intraepithelial neoplasia. BJU Int 2000;85:155–159.

The Prostate 55:55 – 64 (2003)

Molecular Characterization of Prostatic Small-Cell Neuroendocrine Carcinoma

Nigel Clegg,[1] Camari Ferguson,[1] Lawrence D. True,[2] Hugh Arnold,[1] Alec Moorman,[1] Janna E. Quinn,[3] Robert L. Vessella,[3] and Peter S. Nelson[1]*

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, University of Washington Seattle, Washington
[2]Department of Pathology, University of Washington Seattle, Washington
[3]Department of Urology, University of Washington Seattle, Washington

OBJECTIVES. A subset of prostate carcinomas is composed predominantly, even exclusively, of neuroendocrine (NE) cells. In this report, we sought to characterize the gene expression profile of a prostate small cell NE carcinoma by assessing the diversity and abundance of transcripts in the LuCaP 49 prostate small cell carcinoma xenograft.

METHODS. We constructed a cDNA library (PRCA3) from the LuCap 49 prostate small cell xenograft. Single pass DNA sequencing of randomly selected cDNA clones followed by sequence assembly and annotation produced a library of Expressed Sequence Tags (ESTs) representing the LuCaP 49 transcriptome. Comparative sequence analysis with ESTs derived from prostate adenocarcinoma libraries was performed using statistical algorithms designed to identify differentially expressed sequences. Putative NE cell-specific genes were further examined by Northern analysis.

RESULTS. Sequence assembly and analysis identified 1,447 distinct genes expressed in the LuCaP 49 cDNA library. These include cDNAs encoding the NE markers secretogranin (SCG2), CD24, and ENO2. Northern analysis revealed that three additional genes, ASCL1, INA, and SV2B are expressed in LuCaP 49 but not in various prostate cancer cell lines or xenografts. Fifteen genes were identified with a statistical probability ($P > 0.9$) of being up-regulated in LuCaP 49 small cell carcinoma relative to prostate adenocarcinoma (two primary prostate adenocarcinomas and the LNCaP prostate adenocarcinoma cell line).

CONCLUSIONS. Prostate small cell carcinoma expresses a diverse repertoire of genes that reflect characteristics of their NE cell of origin. ASCL1, INA, and SV2B are potential molecular markers for small cell NE tumors and NE cells of the prostate. This small cell NE carcinoma gene expression profile may yield insights into the development, progression, and treatment of subtypes of prostate cancer. *Prostate 55: 55–64, 2003.* © 2003 Wiley-Liss, Inc.

KEY WORDS: **xenograft; cDNA; expressed sequence tag; digital expression; database**

INTRODUCTION

The prostate epithelium is composed of three primary cell types: basal cells, luminal secretory cells, and neuroendocrine (NE) cells. NE cells display hybrid epithelial/neural/endocrine characteristics and have variably prominent dendritic processes [1,2]. Based on ultrastructural studies, two subtypes have been identified, an open subtype with apical processes extending to the glandular lumen and a closed subtype [1,2]. Ultrastructural studies, biochemical analyses, and histochemical staining provide evidence for functionally

Grant sponsor: CaPCURE Foundation; Grant sponsor: Department of Defense; Grant numbers: DAMD 17-98-1-8499, PC991274; Grant sponsor: National Cancer Institute; Grant number: CA75173; Grant sponsor: Department of Veterans Affairs.

*Correspondence to: Peter S. Nelson, M.D., Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024.

Received 27 November 2001; Accepted 1 November 2002
DOI 10.1002/pros.10217

diverse subtypes of NE cells within the prostate [3,4]. These cells secrete a wide range of peptides known to stimulate cell growth (and perhaps cell secretion) in an autocrine and paracrine fashion. However, the role of NE cells in both normal prostate development and in prostate carcinogenesis is poorly understood [5,6].

NE differentiation in prostatic malignancy has been grouped into three categories: (1) focal differentiation of cells with NE features in a conventional adenocarcinoma, (2) carcinoid tumor of the prostate, and (3) small cell undifferentiated NE carcinoma of the prostate [7]. Rare NE cells are found in virtually all prostate adenocarcinomas. Carcinoid tumors are usually foci within a conventional adenocarcinoma; pure prostate carcinoids are extremely rare. Small cell undifferentiated carcinoma is rare, representing only 1–2% of all prostate malignancies. Although small cell undifferentiated carcinoma is most often seen as a component of a conventional adenocarcinoma, the pure small cell tumor has an aggressive course [8,9].

Several model systems for studying the role of NE cells in prostate cancer have been described. One approach has been to express genes with oncogenic potential in mice using heterologous promoters [10–15]. Most of these transgenic models used promoters that do not restrict gene expression to the prostate. Employing a more directed approach, Masumori et al. [15] used the rat probasin promoter to drive expression of the SV40 large T antigen specifically in prostate epithelium. With advancing age, low-grade prostatic intraepithelial neoplasia (PIN), high-grade PIN, microinvasion, invasive carcinoma, and poorly or undifferentiated carcinoma with NE differentiation developed in the prostates in sequential order. Alternatively, studies of human prostate cancers implanted into immune deficient mice have also provided insights into the role of NE cells in prostate carcinogenesis. Androgen deprivation of the prohormone convertase-310 human prostate cancer xenograft induces NE differentiation without proliferation, and may serve as a model for the role of NE cells in hormone refractory prostate cancer [16]. Three androgen-insensitive small cell prostate cancer xenografts have been described (UCRU-PR-2, WISH-PC2, and LuCaP 49) that are capable of proliferative growth [17–20]. These xenografts are composed of actively dividing NE-like cells that express a variety of NE-enriched molecular markers, but otherwise, little is known about the genes that they express.

LuCaP 49 is a xenograft that exhibits a rapid growth rate (doubling time 6.5 days) and is composed almost exclusively of cells with a NE/small cell carcinoma phenotype [19]. As such, LuCaP49 provides a rare opportunity to study the repertoire of genes expressed in NE-like cells of the prostate. Here, we report the isolation and characterization of 2,096 Expressed Sequence Tags (ESTs) from a LuCaP 49 cDNA library that identifies 1,447 distinct genes. Together, these sequences represent a partial transcriptome reflecting the diversity and relative abundance of the genes and their cognate transcripts that are expressed in prostate small cell carcinoma. Many of these genes are expressed in other cell types, but several are highly enriched in NE cells. In addition, a statistical analysis of EST frequencies was used to identify genes that are expressed at higher levels in LuCaP 49 than in primary prostate adenocarcinomas and in the LNCaP prostate adenocarcinoma cell line. These expression differences may reflect unique features of small cell prostate cancers that can further the understanding of the role of NE cells in the development of small cell carcinoma and adenocarcinoma of the prostate.

MATERIALS AND METHODS

The establishment and characterization of the LuCaP 49 NE small cell xenograft is described in detail elsewhere [19]. LuCaP 49 was derived from an omental mass removed during surgery. The tumor was isolated from a 71-year-old male originally diagnosed with clinical stage B-II prostate carcinoma 4 years prior to obtaining the tumor. The xenograft was established in Fox Chase CB.17 SCID mice (Charles River Laboratories, Wilmington, MA) and has been serially passaged for 5 years. Histological analysis of sections adjacent to flash-frozen tissue revealed predominantly NE cells interspersed with approximately 5% mouse stromal cells.

PolyA+RNA was isolated from a LuCaP 49 xenograft sample using Trizol reagent (Life Technologies, Carlsbad, CA) and oligo-dT columns (Life Technologies, Carlsbad, CA). A cDNA library designated PRCA3 was constructed in the pSPORT vector according to protocols we have previously described [21]. DNA sequencing and Northern blot analysis was performed by standard methods as described in Clegg et al. [22]. Details of the PRCA3 library construction are also available at http://www.pedb.org. The average insert size of PRCA3 cDNA clones is 1.2 kb.

DNA sequences were stored, clustered, and annotated using the Prostate Expression Database (PEDB) and associated data analysis tools [23,24]. Briefly, vector, *E. coli* and interspersed repeats were masked in ESTs using Cross_Match (bozeman.mbt.washington.edu/phrap.doc/general.html) and RepeatMasker (ftp.genome.edu/RM/RepeatMasker.html). Phrap (P. Green, University of Washington, Washington), which incorporates estimates of sequence quality, was used to cluster the masked sequences and generate a consensus sequence for each assembly. Each distinct cluster was annotated by searching Unigene

[25], GenBank [26], or dbEST [27] using BLASTN (http://blast.wustl.edu). Annotations were assigned using a Perl-based script to select the lowest *P* value where the maximum *P* value was e-20 and the minimum BLAST score was 300. The biological role for each species was assigned using the categories described in Adams et al. [28].

SAGE libraries SAGE_PR317_prostate_tumor and SAGE_PR317_normal_prostate are listed at the NCBI Library Browser web site (www.ncbi.nlm.nih.gov/SAGE/sagelb.cgi); expression tags were downloaded from SAGEmap's anonymous FTP site (ftp://ncbi.nlm.nih.gov/pub/sage/seq/). EST sequences for lung cancer libraries NCI_CGAP_Lu5, NCI_CGAP_Lu6, NCI_CGAP_Lu24, NCI_CGAP_Pr3, and NCI_CGAP_Pr12 are listed at http://cgap.nci.nih.gov/Tissues/LibraryFinder. Pr12 and Pr3 sequences are also stored in the PEDB. Comparisons between sequence tags (ESTs or SAGE tags) were performed using the method of Audic and Claverie [29], which allows statistical analysis of small samples of expression tags.

RESULTS

The prostate small cell cancer xenograft LuCaP 49 was used to construct a cDNA library (PRCA3). Morphological and immunohistochemical data show that LuCaP 49 is nearly identical to the primary tumor from which it was derived, hence it is a good source of material for the study of NE gene expression in the prostate [19]. In brief, both the xenograft and the primary tumor are composed of undifferentiated cells characterized by a high nuclear:cytoplasmic ratio, and nuclei with a fine heterochromatin pattern and inconspicuous nucleoli (Fig. 1). Both the original tumor and the xenograft express the NE markers synaptophysin and neuron specific enolase in similar numbers of cells (40–80% and 80%, respectively); and neither expresses the adenocarcinoma markers PSA and the androgen receptor. One difference between the primary cancer and the xenograft is that fewer xenograft cells express the NE marker Chromogranin A (Fig. 1).

Clones from the PRCA3 library were randomly selected and partially sequenced to generate 2,096 high quality ESTs representing a partial transcriptome of this tissue. ESTs were assembled using the Phrap sequence assembly program to produce 1,577 clusters. Each cluster was used to query the Unigene, nonredundant Genbank, and dbEST databases. Based on shared annotations, the clusters were further consolidated and assigned to 1,447 distinct transcripts. Twenty-four transcripts were of murine origin. These presumably represent tissue contamination from the xenograft host. For classification purposes, the remaining 1,423 are assumed to be of human origin; however, 148 transcripts were not homologous to any known

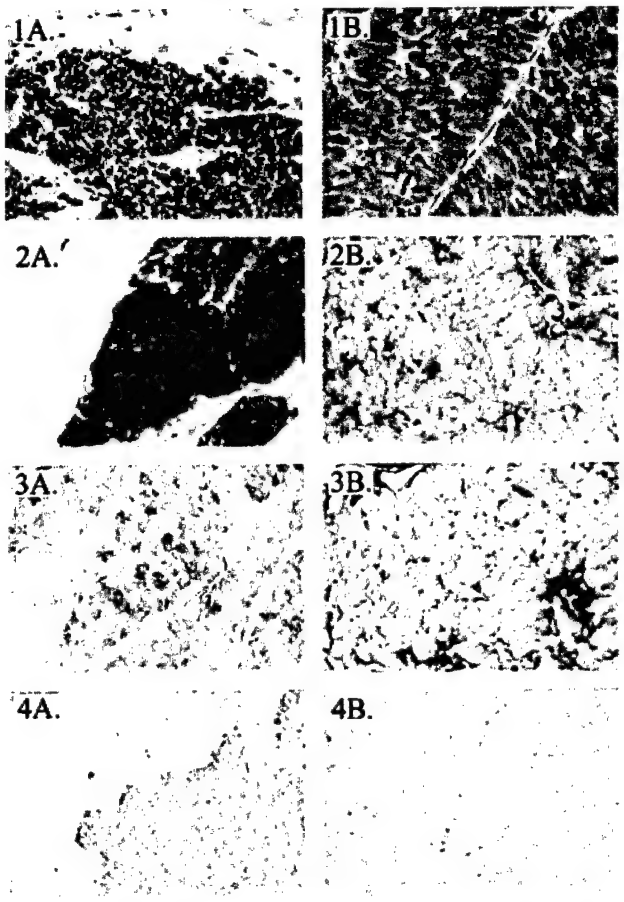


Fig. 1. Phenotypes of a prostate small cell carcinoma (**left panel**) and the LuCaP 49 xenograft that was derived from it (**right panel**). **1A**: Primary tumor. Sheet of undifferentiated carcinoma cells invading omental fat. The tumors have a high nuclear:cytoplasmic ratio, a fine heterochromatin pattern, and inconspicuous nucleoli. **1B**: Xenograft. Aggregate of cohesive, undifferentiated carcinoma cells with histologic features similar to 1A, and frequent mitoses. [1A, 1B: hematoxylin and eosin, original magnification 400×]. **2A**: Tumor. Uniform, intense, synaptophysin immunoreactivity. **2B**: Xenograft. Variably intense, cytoplasmic synaptophysin immunoreactivity. **3A**: Tumor. Focal cytoplasmic expression of chromogranin A. **3B**: Xenograft. Chromogranin A expression in the periphery of the cytoplasm of a minority of tumor cells. **4A,B**: Negative controls with no immunoreactivity. [All immunostains: Ni.DAB black, cytoplasmic reaction product; faint grey, nuclear green counterstain; 400× magnification].

nucleotide or protein sequence recorded in the public databases, and an undetermined number of these could be from murine mRNAs. Alternatively, the unannotated species may represent novel human transcripts.

A complete summary of all isolated transcripts representing both known and uncharacterized genes can be found at http://www.pedb.org. Mitochondrial sequences represent 4.3% of all the ESTs, but were counted as a single species. The majority of the nonmitochondrial genes are represented by a single EST

(1141; 80%), which is expected because most tissues contain 15,000–30,000 different transcript types [30], and our study sampled 2,096 ESTs. Only 13 (<1%) of all genes were represented by more than six ESTs. When the genes were assigned functional roles, 3 of the 13 most abundant transcripts were ribosomal, 3 were cytoskeletal, and the remainders were of diverse function. The most frequently sampled transcript encodes translation elongation factor 1 alpha 1, represented by 31 ESTs.

The general distribution of biological roles in the NE PRCA3 sample (Table I) is similar to those observed for both LNCaP prostate adenocarcinoma cells and the normal prostate [21,22]. A high proportion of all transcripts (56%) could not be assigned any functional role. If the unclassified species are excluded, the functional category of gene expression comprised the greatest number of transcripts (38%) and cell signaling the second (20%). The gross similarity to LNCaP cells may reflect the secretory nature of both the epithelial and NE cells.

Potential Markers for NE-Like Cells

A variety of molecular markers have been described for NE cells and small cell tumors of the prostate. Among these are members of the granin family of acidic glycoproteins, which are quantitatively the major components of dense-core secretory granules and which are required for the regulated secretion of prohormones [31,32]. ESTs for secretogranins 1, 2, and 3 (CHGB, secretogranin (SCG2), and SCG3) were detected in the PRCA3/LuCaP 49 library (Table II). ESTs for chromogranin A, a key regulator of dense-core secretory biogenesis [32], were not isolated; however, chromogranin A is histochemically detectable in the LuCaP 49 xenograft (Fig. 1; [19]). An EST was also detected that encodes NE secretory protein 55 (NESP55), a chromogranin-like protein that may function in secretion [33].

TABLE I. Roles of LuCaP 49/PRCA3 Species

Category	ESTs	Species
Gene expression	424 (0.202)[a]	231 (0.381)
Metabolism	162 (0.077)	94 (0.155)
Cell division	66 (0.032)	45 (0.074)
Cell signaling	164 (0.079)	121 (0.200)
Defense	106 (0.051)	66 (0.109)
Cell structure	100 (0.048)	49 (0.081)
Unclassified	949 (0.459)	816 (—)
Mitochondrial	90 (0.043)	1 (—)
Mouse	35 (0.017)	24 (—)
Total	2,096	1,447

[a]Proportion of total.

ESTs were identified that encode several other known NE-enriched markers. Enolase 2 gamma (ENO2), the CD24 antigen (CD24), and synaptophysin (SYP) are standard histochemical markers in the prostate. Our analysis identified two other potential markers of prostate NE cells, the achaete-scute complex (Drosophila) homolog-like 1 (ASCL1), and secretagogin (SECRET). ASCL1 encodes a basic-helix-loop-helix transcription factor that is highly expressed in medullary thyroid cancers and lung tumors with NE properties [34,35]. SECRET encodes a calcium binding protein that is enriched in some NE cells and may function in cell proliferation [34].

The origin of small cell cancers and adenocarcinomas of the prostate is an area of current debate and it remains uncertain to what extent gene expression profiles in the two cancer types may overlap [36–39]. To investigate whether the 'NE-enriched' transcripts in PRCA3 are also expressed and abundant in adenocarcinomas or normal prostate, we compared EST frequencies in the PRCA3 library to expression tag frequencies in the PR317 and LNCaP SAGE libraries available at the National Cancer Institute SAGE website. The original PR317 tumor was a primary adenocarcinoma of the prostate while the LNCaP cell line is derived from a metastatic adenocarcinoma. These SAGE libraries comprise 65109 and 22637 sequence tags respectively, and thus represent a deep sampling of the transcriptomes expressed by these cell and tissue types. Using the probability function of Audic and Claverie [29], 8 of the 9 potential NE markers had a high likelihood ($P > 0.9$) of elevated expression in PRCA3/LuCaP 49 relative to the other samples (Table II). Only NESP55 had a P-value less than 0.9. Five of the 9 potential markers (SYP, SCG2, SCG3, CHGB, and ASCL1) were represented exclusively in the PRCA3/LuCaP 49 library sample; and one (SECRET) was represented only in the PRCA3/LuCaP 49 NE and PR317 adenocarcinoma library samples. Surprisingly, CD24 and ENO2 ESTs were found in all of the libraries. Since CD24 and ENO2 are used as markers for prostate NE cells, this finding demonstrates the importance of using statistical methods to evaluate expression profiles instead of relying exclusively on the presence or absence of ESTs.

To confirm the statistical observations of Table II, we examined the expression of three transcripts, SCG2, ASCL1, and SECRET, in a variety of cancer cell lines using Northern analysis (Fig. 2). SCG2 RNA is expressed in the LuCaP 49 small cell carcinoma, but is not detectable in the adenocarcinoma xenograft LuCaP 73; nor is it detectable in whole normal prostate tissue or the cell lines DU145, LNCaP, and PC3. Like SCG2, the ASCL1 gene is expressed exclusively in LuCaP 49. SECRET is expressed in all of the samples, but it is

TABLE II. Neuroendocrine (NE) and Neural Genes Expressed in LuCaP 49

			cDNA library (tags per million)					
		Unigene	Known prostate NE	PRCA3[a] small cell	PR317 normal[b]	Tumor[c]	LNCaP[d] cell line	Lowest probability of differential expression[e]
Neuroendocrine								
ASCL1	Achaete–scute complex (*Drosophila*) homolog-like 1	Hs.1619	–	954	0	0	0	$0.998 < P < 0.999$
CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen)	Hs.286124	+	2,386	117	307	176	$0.999 < P < 1.000$
ENO2	Enolase 2, gamma	Hs.146580	+	477	50	15	44	$0.95 < P < 0.96$
NESP55	NE secretory protein 55	Hs.113368	–	477	302	230	132	$0.70 < P < 0.80$
CHGB	Secretogranin 1 (chromogranin B)	Hs.2281	+	477	0	0	0	$0.98 < P < 0.99$
SCG2	Secretogranin 2 (chromogranin C)	Hs.75426	+	1,908	0	0	0	$0.999 < P < 1.00$
SCG3	Secretogranin 3	Hs.22215	+	477	0	0	0	$0.98 < P < 0.99$
SECRET	Secretagogin	Hs.116428	–	477	0	61	0	$0.97 < P < 0.98$
SYP	Synaptophysin	Hs.75667	+	954	0	0	0	$0.998 < P < 0.999$
Neural								
INA	Internexin neuronal intermediate filament protein, alpha	Hs.76888	–	477	0	0	0	$0.98 < P < 0.99$
KAL1	Kallmann syndrome 1	Hs.89591	–	477	0	0	0	$0.98 < P < 0.99$
SV2B	Synaptic vesicle protein 2B homolog	Hs.194035	–	477	16	30	0	$0.98 < P < 0.99$
NLGN3	Neuroligin 3	Hs.47320	–	477	65	15	0	$0.96 < P < 0.97$
SYT13	Synaptogamin 13	Hs.12365	–	477	0	0	0	$0.98 < P < 0.99$
TPH	Tryptophan hydroxylase	Hs.144563	–	477	0	0	0	$0.98 < P < 0.99$
RTN3	Reticulon 3	Hs.252831	–	477	49	76	441	$0.40 < P < 0.50$
RTN4	Reticulon 4 (foocen, ASY, NOGO)	Hs.65450	+	477	251	122	44	$0.70 < P < 0.80$
ROBO1	Roundabout	Hs.301198	–	1,908	0	15	0	$0.99 < P < 1.00$

[a]PRCA3 cDNA library from small cell xenograft LuCaP49; 2,069 ESTs.
[b]PR317 normal prostate tissue 59419 SAGE tags.
[c]PR317 adenocarcinoma 65109 SAGE tags.
[d]LNCaP cell line 22637 SAGE tags.
[e]PRCA3 vs. PR317 and LNCaP; method of Audic and Claverie [29].

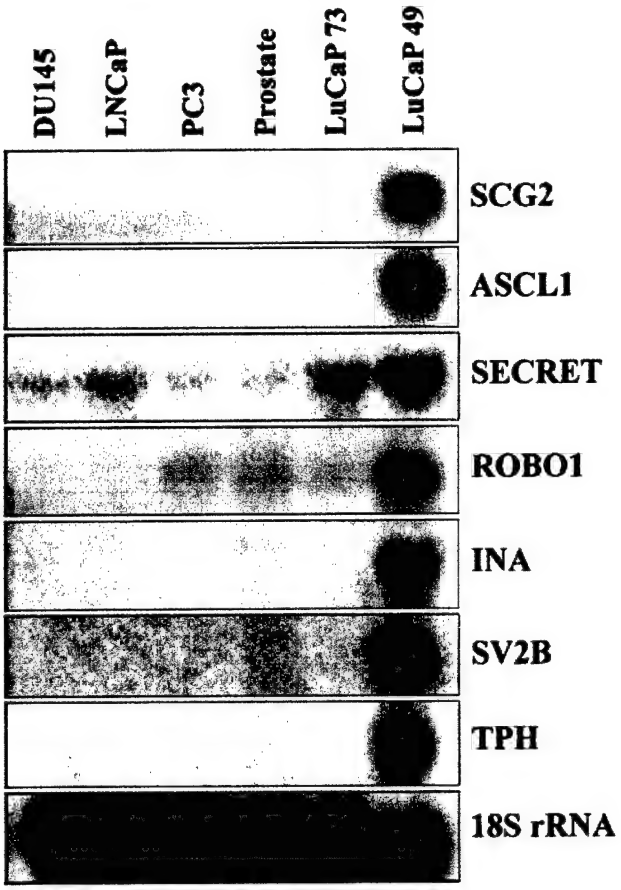


Fig. 2. LuCaP 49-enriched transcripts. Northern blots of total RNA from cell lines (DU145, LNCaP, PC3), an adenocarcinoma xenograft (LuCaP 73) [20], and LuCaP 49. SCG2, secretogranin 2; ASCL1, achaete-scute complex-like 1 (*Drosophila*); SECRET, secretagogin; ROBO1, roundabout, axon guidance receptor, homolog 1 (*Drosophila*); INA, internexin neuronal intermediate filament protein alpha; SV2B, synaptic vesicle protein 2B homolog; TPH, tryptophan hydroxylase.

enriched approximately 4-fold in LuCaP 49 RNA relative to the other samples. These data confirm the presence of SCG2 transcripts in LuCaP 49 and identify ASCL1 mRNA as a new potential marker for NE-like cells of the prostate. Despite SECRET's enrichment in LuCaP 49, it's presence in other cell-types make it less attractive as a marker.

Since NE and neural cells share many characteristics, we also searched for genes in the PRCA3 library that are enriched in neural tissues. Internexin neuronal intermediate filament protein alpha (INA) is found predominantly in the CNS [40], but a BLAST search of the NCI_CGAP_Lu5 and NCI_CGAP_Lu6 libraries revealed it is also expressed in lung cancers with NE characteristics. Reticulon 3 and 4 transcripts are enriched in the brain and in several other tissues [41–43]. Synaptotagmin 13 (SYT13) belongs to a family of calcium-binding synaptic vesicle proteins [44], while the function of the synaptic vesicle protein 2B homolog (KIAA0737) is unknown. The Kallmann Syndrome 1 (KAL1) and Roundabout homolog 1 (ROBO1) gene products may act in axon guidance [45–47]. Finally, the tryptophan dehydroxylase gene (TPH) encodes an enzyme that catalyses the rate limiting step in serotonin synthesis. Serotonin production is well-documented in prostate NE cells [3,4]. Any of these markers might serve to differentiate Ne-like cells from other cell types in the prostate.

With the exception of RTN3 and RTN4, all of the neural genes identified in Table II are predicted to be differentially expressed when PRCA3/LuCap 49 EST frequencies are compared to those in the PR317 and LNCaP adenocarcinoma libraries. We examined the expression of 6 of the genes (RTN4, NLGN3, ROBO1, INA, SV2B, and TPH) in the cell lines and xenografts described previously. Both RTN4 and NLGN3 were expressed in all of the cell lines tested (data not shown). The ROBO1 gene is unique in being expressed in a subset of cell lines. Furthermore, it is only enriched 3 fold in LuCap 49 relative to whole prostate tissue. In contrast, INA, SV2B and TPH transcripts are present in the LuCaP 49 xenograft and are not detected in the LNCaP, PC3, and DU145 cell lines, all of which are derived from adenocarcinoma metastases (Fig. 2). INA also expresses a transcript that is detected in all cell types tested (data not shown). Hence, we have identified three genes with transcripts that are expressed in LuCap 49, but not in adenocarcinoma-derived samples.

Virtual Expression Analysis of Prostate NE Small Cell Carcinoma

An alternate strategy for finding genes that are relevant to small cell cancer and NE cell biology is to identify transcripts that are highly expressed in NE-like cells relative to other tumor types. This approach does not require any *a priori* knowledge of gene function. We restricted our analysis to transcripts identified in PRCA3/LuCaP 49 that are also expressed in libraries derived from other NE cancers: NCI_CGAP_Lu24 (36609 ESTs), NCI_CGAP_Lu5 (20359 ESTs), or NCI_CGAP_Lu6 (209 ESTs). Lu24 and Lu5 were constructed from NE lung carcinoid tumors; Lu6 was made from a small cell lung carcinoma. Five hundred and ninety-one PRCA3 species were represented in one or more of these cancer libraries. While most of these NE-expressed species are unlikely to represent 'NE-specific' genes, their expression profiles may vary in other cancer types.

We compared the EST frequencies of the 591 NE-expressed transcripts in PRCA3 with the EST frequencies of the same genes in NCI_CGAP_Pr3, a primary adenocarcinoma of the prostate. Using statistical methods [29], 132 of 591 species were predicted to

have significantly different levels of expression in the PRCA3 small cell carcinoma library relative to the Pr3 adenocarcinoma library ($P > 0.9$). Another round of selection was performed by comparing EST frequencies from PRCA3 and a library derived from a prostate cancer adenocarcinoma bone metastasis, NCI_CGAP_Pr12. Forty NE-expressed genes were differentially expressed ($P > 0.9$). A list of these genes is posted at http://www.pedb.org. Since many of the genes have housekeeping functions and may simply reflect differences in metabolic activity, one final round of statistical selection was applied. EST frequencies of the 40 PRCA3 genes were compared to EST frequencies from the LNCaP cell line. This comparison identified 15 genes with a high probability of differential expression ($P > 0.90$; Table III).

The CD24 antigen and SCG2 genes are predicted to be more highly expressed in prostate NE small cell cancer than in normal prostate, prostate adenocarcinomas, and the LNCaP cell line (Table III), as are the nearly ubiquitously expressed tubulin genes TUBA3 and TUBB. Three other differentially expressed genes are of particular interest with respect to cancer. ALL fused gene from chromosome 1 (AF1Q) is both highly expressed in the thymus and is fused with a variety of other genes in leukemias [48]. Anti-apoptotic-like after growth factor withdrawal (API5L1) is a gene that may protect cells from apoptosis [49], and retinoblastoma binding protein 7 (RBBP7) is found in histone deacetylation complexes and interacts with the BRCA1 protein [50,51].

Other Genes of Interest

NE and malignant NE cells of the prostate have been reported to synthesize and secrete a variety of neuropeptides, including members of the calcitonin gene family, gastrin-releasing-peptide, somatostatin, alpha-human chorionic gonadotropin, thyroid-stimulating hormone (TSH)-like peptide and parathyroid hormone-related protein. Our sample of over 2000 PRCA3 ESTs did not include transcripts from these genes; however, other neuropeptides were found, including calcitonin gene-related peptide-receptor component protein (GCRP-RPC), thyroid-hormone receptor interactor 7 (TRIP7), thyroid receptor interacting protein 15 (TRIP15), and thyroid hormone binding protein p55 (P4HB).

Another area of active research is the role of apoptosis in prostate cancers. ESTs were isolated for programmed cell death 6-interacting protein (PDCD6IP), nerve growth factor receptor (TNFRSF16) associated protein 1 (NGFRAP1), API5-like 1 (API5L1), myeloid-cell leukemia sequence 1, BCL2 related (MCL1), Tax1 binding protein 1 (TAX1BP1), programmed cell death 4 (PDCD4), TGFB1 induced anti-apoptotic factor 1 (TIAF1), apoptosis antagonizing transcription factor

TABLE III. Genes Highly Expressed in LuCaP 49 Relative to Adenocarcinoma-Derived Samples

			cDNA library (tags per million)				
Gene	Description	Unigene Id	PRCA3[a]	PR3[b]	Pr12[c]	LNCaP[d]	Lowest P value[e]
HLA-A	MHC class I-A	Hs.181244	3,817	231	0	0	$0.998 < P < 0.999$
RBBP7	Retinoblastoma binding-protein 7	Hs.31314	3,340	0	0	560	$0.95 < P < 0.96$
SFRS3	Splicing factor arginine/serine rich 3	Hs.167460	2,386	0	0	187	$0.94 < P < 0.95$
CD24	CD24 antigen	Hs.286124	2,386	0	0	0	$0.98 < P < 0.99$
LDHA	Lactate dehydrogenase A	Hs. 2795	2,386	0	0	187	$0.94 < P < 0.95$
AF1Q	ALL1 fused gene from chromosome q1	Hs.75823	1,908	0	0	0	$0.96 < P < 0.97$
SCG2	SCG2	Hs.75426	1,908	0	0	0	$0.96 < P < 0.97$
API5L1	API5-like 1	Hs.227913	1,908	0	0	0	$0.96 < P < 0.97$
	ESTs similar to mucin 2 precursor	Hs.111911	1,431	0	0	0	$0.92 < P < 0.93$
LAPTM4A	Lysosomal associated protein transmembrane 4 alpha	Hs.111894	1,431	0	0	0	$0.92 < P < 0.93$
POLR2H	Polymerase (RNA) II (DNA-directed) polypeptide H	Hs.3128	1,431	0	0	0	$0.92 < P < 0.93$
FLJ20160	Hypothetical protein FLJ20160	Hs.23412	1,431	0	0	0	$0.92 < P < 0.93$
RACGAP1	GTPase activating protein	Hs.23900	1,431	0	0	0	$0.92 < P < 0.93$
TUBA3	Tubulin alpha, brain specific	Hs.272897	6,679	0	0	373	$0.999 < P < 1.00$
TUBB	Tubulin beta polypeptide	Hs.179661	4,771	231	1,142	746	$0.98 < P < 0.99$

[a]PRCA3 2,096 ESTs.
[b]PR3 4,325 ESTs.
[c]Pr12 3,500 ESTs.
[d]LNCaP 5,362 ESTs.
[e]Lowest probability of differential expression among three comparisons.

(DED), and BCL2-like 1 (BCL2L1). The LuCaP 49 xenograft contains foci of necrosis, which may account for the large number of expressed anti- and pro-apoptotic genes [19].

DISCUSSION

The role of NE cells in normal prostate development and in the etiology and progression of both adenocarcinoma and small-cell cancers of the prostate is, at present, unclear. To help address these broad issues we have created a cDNA library, PRCA3, from the LuCaP 49 small cell xenograft and have begun characterizing the transcriptome expressed by this tumor type. This report summarizes the identities of 2,096 cDNA clones derived from the PRCA3 library. The sequences assemble into 1,447 distinct transcripts, some of which are highly enriched in NE and NE-like cells. We confirmed the presence of several known, NE markers in LuCaP 49 (ASCL1, CD24, ENO2, CHGB, SCG2, SCG3, and SYP). Eighteen other genes including INA, SV2B, and TPH were found to be highly expressed in LuCaP 49 relative to normal prostate or cell lines derived from prostate adenocarcinomas.

A fundamental concern in describing NE cells in prostate cancer is the extent to which a malignant cell may be considered a NE cell. The dual expression of epithelial markers (such as PSA) and NE markers (such as chromogranin A) has been demonstrated in some cancer cells [3,4]. Similarly, two cell lines derived from adenocarcinomas can be induced to express NE markers [52–54], and chromogranin A is expressed at low frequency in LNCaP cells (http://pedb.org). These phenomena may reflect the ontogeny of prostate cancers. One hypothesis is that basal cells, secretory luminal cells, and NE cells arise from a pluripotent stem cell [38]. Hence, any prostate cancer may be able to express a variety of the markers typically associated with the terminal differentiation of these cell types [38]. Nevertheless, enrichment of a molecular marker in a small cell tumor represents the first step towards identifying potential determinants that will serve to characterize the functional attributes of both NE-like cancers and true NE cells.

The characterization of transcripts expressed in the PRCA3 library revealed nine NE-enriched genes, including one, ASCL1, which had not previously been observed in the prostate. Depletion of ASCL1 transcripts in cultured small cell lung cancers decreases the expression of NE markers [35]. Further, disruption of the mouse homolog of ASCL1, MASH1, prevents NE differentiation in the lung but not in the gut or pancreas [35]. These data suggest that ASCL1 participates in NE cell differentiation, either by affecting cell fate or by modulating the expression of factors important for terminal differentiation. MASH1 null mice die with 24 hr of birth, approximately 2 months before NE cells are detectable in the prostate. Consequently, the role of MASH1 in prostate development has not been described. If ASCL1 plays a role in mediating prostate cellular differentiation, then expressing ASCL1 in non-NE human prostate cancer cell lines may provide insights into the genes it regulates. The cell lines LNCaP and C4-2 may be especially informative as both are competent to express NE markers upon treatment with a variety of physiological and pharmacological agents [52–54]. Alternatively, a transgene under the control of a prostate specific promoter, such as the probasin promoter, make it possible to address this question. A recently described mouse model system for the induction of prostate small cell cancers should permit further investigation of ASCL1's role in progression towards small cell prostate cancer [15].

Eight neural genes were identified among the 1,447 gene species in the PRCA3 library. Three of the six genes studied by Northern analysis were expressed only in LuCaP 49 and not in other non-NE xenografts or prostate cancer cell lines. This suggests that they may be good markers for NE-like cells. Tests with a large sample of primary cancers and normal tissues will be needed to confirm this preliminary conclusion, as primary cancers and cancer cell lines may exhibit diverse gene expression profiles. For example, in our study the gene-encoding Roundabout was expressed in LuCaP 49 and PC3 cells but was not expressed at detectable levels in LNCaP or DU145 cell lines.

In addition to NE markers, 15 genes are expressed at significantly higher levels in PRCA3/LuCaP49 compared to libraries of adenocarcinoma origin. CD24 and SCG2 are expressed in a tissue specific manner, but all of the other genes were found to be expressed in adenocarcinomas of the prostate or in LNCaP cells (see http://www.ncbi.nlm.nih.gov/SAGE/SAGEcid.cgi/). The most highly represented species are the two tubulins TUBA3 and TUBB. The high level of tubulin mRNA may make LuCaP 49 cells particularly susceptible to taxane chemotherapuetic drugs, which induce apoptosis by promoting microtubule polymerization [55]. Taxanes may also act to impair function of anti-apoptotic members of the BCL family of proteins. Expression of bcl-xL, which we observed in PRCA3, is reduced in some cell lines in response to docetaxel. Thus, LuCaP 49 may serve as a good model system for the role of this class of chemotherapeutic drugs on small cell cancers of the prostate.

ACKNOWLEDGMENTS

This work was supported in part by the CaPCURE Foundation, grants DAMD 17-98-1-8499 and PC991274

(to PSN) from the Department of Defense, grant CA75173 (to PSN) from the National Cancer Institute, and a grant from the Department of Veterans Affairs (to RLV). We thank D. Abbott for expert help with the figures.

REFERENCES

1. di Sant'Agnese P, de Mesy Jensen K. Endocrine-paracrine cells of the prostate and prostatic urethra; an ultrastructural study. Hum Pathol 1984;15:1034–1041.
2. di Sant'Agnese P, de Mesy Jensen K. Human prostatic endocrine-paracrine (APUD) cells: Distributional analysis with a comparison of serotonin and neuron specific enolase immunoreactivity and silver stains. Arch Pathol Lab Med 1985;109: 607–612.
3. di Sant'Agnese PA. Neuroendocrine differentiation in human prostatic carcinoma. Hum Pathol 1992;23:287–296.
4. di Sant'Agnese PA. Neuroendocrine cells of the prostate and neuroendocrine differentiation in prostatic carcinoma: A review of morphologic aspects. Urology 1998;51(Suppl 5A):121–124.
5. Bonkhoff H. Neuroendocrine cells in benign and malignant prostate tissue: Morphogenesis, proliferation, and androgen receptor status. The Prostate 1998;36(Suppl 8):18–22.
6. Abrahamsson PA. Neuroendocrine cells in tumor growth of the prostate. Endocrine-Related Cancer 1999;6:503–519.
7. di Sant'Agnes PA, Cockett ATK. Neuroendocrine differentiation in prostatic malignancy. Cancer 1996;78:357–361.
8. Cohen RJ, Glezerson G, Haffejee Z. Prostatic carcinoma. Histological and immuno-histological factors affecting prognosis. Br J Urol 1991;66:405–410.
9. di Sant'Agnese PA. Neuroendocrine differentiation in carcinoma of the prostate: Diagnostic, prognostic, and therapeutic implications. Cancer 1992;70:254–258.
10. Skalnik DG, Dorfman DM, Williams DA, Orkin SH. Restriction of neuroblastoma to the prostate gland in transgenic mice. Mol Cell Biol 1991;11:4518–4527.
11. Perez-Stable C, Altman NH, Brown J, Harbison M, Cray C, Roos BA. Prostate, adrenocortical, and brown adipose tumors in fetal globin/T antigen transgenic mice. Lab Invest 1996;74: 363–373.
12. Perez-Stable C, Altman NH, Mehta PP, Deftos LJ, Roos BA. Prostate cancer progression, metastasis, and gene expression in transgenic mice. Cancer Res 1997;57:900–906.
13. Garabedian EM, Humphrey PA, Gordon JI. A transgenic mouse model of metastatic prostate cancer originating from neuroendocrine cells. Proc Natl Acad Sci USA 1998;95:15382–15387.
14. DiGiovanni J, Kiguchi K, Frijhoff A, Wilker E, Bol DK, Beltran L, Moats S, Ramirez A, Jorcano J, Conti C. Deregulated expression of insulin-like growth factor 1 in prostate epithelium leads to neoplasia in transgenic mice. Proc Natl Acad Sci USA 2000;97: 3455–3460.
15. Masumori N, Thomas TZ, Chaurand P, Case T, Paul M, Kasper S, Caprioli RM, Tsukamoto T, Shappell SB, Matusik RJ. A probasin-large T antigen transgenic mouse line develops prostate adenocarcinoma and neuroendocrine carcinoma with metastatic potential. Cancer Res 2001;61:2239–2249.
16. Jongsma J, Oomen MH, Noordzij MA, Van Weerden WM, Martens GJ, van der Kwast TH, Schroder FH, van Steenbrugge GJ. Androgen deprivation of the prohormone convertase-310 human prostate cancer model system induces neuroendocrine differentiation. Cancer Res 2000;60:741–748.
17. Jelbart ME, Russell PJ, Fullerton M. Ectopic hormone production by a prostatic small-cell carcinoma xenograft line. Mol Cell Endocrinol 1988;55:161–171.
18. Pinthus JH, Waks T, Schindler DG, Harmelin A, Said JW, Bellegrun A, Ramon J, Eshhar Z. WISH-PC2: A unique xenograft model of human prostatic small cell carcinoma. Cancer Res 2000;60:6563–6567.
19. True LD, Buhler K, Quinn J, Williams E, Nelson P, Clegg N, Macoska J, Norwood T, Liu A, Ellis W, Lange P, Vessella R. A neuroendocrine/small cell prostate carcinoma xenograft: LuCaP 49. Am J Pathol 2002;705–715.
20. Buhler K, Quinn J, Whitney S, Lin D, Ellis W, Corey E, Liu A, True L, Lange P, Vessella R. The LuCaP series of human prostate cancer xenografts. J Urol 1999;161:58.
21. Nelson PS, Ng WL, Schummer M, True LD, Liu AY, Bumgarner RE, Ferguson C, Dimak A, Hood L. An expressed-sequence-tag database of the human prostate: Sequence analysis of 1168 cDNA clones. Genomics 1998;47:12–25.
22. Clegg N, Eroglu B, Ferguson C, Arnold A, Moorman A, Nelson PS. Digital expression profiles of the androgen response program. J Steroid Biochem Mol Biol 2002;80:13–23.
23. Hawkins V, Doll D, Bumgarner R, Smith T, Abajian C, Hood L, Nelson PS. PEDB: The Prostate Expression Database. Nucleic Acids Res 1999;27:204–208.
24. Nelson PS, Pritchard C, Eroglu B, Abott D, Clegg N. The human (PEDB) and mouse (mPEDB) prostate expression databases. Nucleic Acids Res 2002;30:218–220.
25. Schuler GD. Pieces of the puzzle: Expressed sequence tags and the catalog of human genes. J Mol Med 1997;75:694–698.
26. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. Nucleic Acids Res 2000;28:15–18.
27. Boguski MS, Lowe TM, Tolstoshev CM. dbEST–database for expressed sequence tags. Nat Genet 1993;4:332–333.
28. Adams MD, Kerlavage AR, Fleischman RD. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 1995; 377(Suppl 28):3–174.
29. Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res 1997;7:986–995.
30. Hastie ND, Bishop JO. The expression of three abundance classes of messenger RNA in mouse tissue. Cell 1976;9:761–774.
31. Tooze SA, Martens GJM, Huttner WB. Secretory granule biogenesis: Rafting to the SNARE. Trends Cell Biol 2001;11: 116–122.
32. Kim T, Tao-Cheng J-H, Eiden LE, Loh YP. Chromogranin A, an On/Off switch controlling dense-core secretory granule biogenesis. Cell 2001;106:499–509.
33. Weiss U, Ischia R, Eder S, Lovisetti-Scamihorn P, Bauer R, Fischer-Colbrie R. Neuroendocrine secretory protein 55 (NESP55): Alternative splicing onto transcripts of the GNAS gene and posttranslational processing of a maternally expressed protein. Neuroendocrinology 2000;71:177–186.
34. Ball DW, Azzoli CG, Baylin SB, Chi D, Dou S, Donis-Keller H, Cumaraswamy A, Borges M, Nelkin BD. Identification of a human achaete-scute homolog highly expressed in neuroendocrine tumors. Proc Natl Acad Sci USA 1993;90:5648–5652.
35. Borges M, Linnoila RI, van de Velde HJK, Chen H, Nelkin BD, Mabry M, Baylin SB, Ball DW. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature 1997;386:852–855.

36. Bonkhoff H, Remberger K. Differentiation pathways and histogenetic aspects of normal and abnormal prostatic growth: A stem cell model. The Prostate 1996;28:98–106.
37. Xue Y, Smedts F, Verhofstad A, Debruyne F, de la Rosette J, Schalken J. Cell kinetics of prostate exocrine neuroendocrine epithelium and their differential interrelationship: New perspectives. The Prostate 1998;36(Suppl 8):62–73.
38. de Marzo AM, Nleson WG, Meeker AK, Coffey DS. Stem cell features of benign and malignant prostate epithelial cells. J Urol 1998;160:2381–2392.
39. Aumuller G, Leonhardt M, Janssen M, Konrad L, Bjartell A, Abrahamsson PA. Neurogenic origin of human prostate endocrine cells. Urology 1999;53:1041–1048.
40. Chan SO, Chiu FC. Cloning and developmental expression of human 66 kD neurofilament protein. Brain Res Mol Brain Res 1995;29:177–184.
41. Moreira EF, Jaworski CJ, Rodriguez IR. Cloning of a novel member of the reticulon gene family (RTN3): Gene structure and chromosomal localization to 11q13. Genomics 1999;58:73–81.
42. Chen MS, Huber AB, van der Haar ME, Frank M, Schnell L, Spillmann AA, Christ F, Schwab ME. Nogo-A is a myelin-associated neurite outgrowth inhibitor and an antigen for monoclonal antibody IN-1. Nature 2000;403:434–439.
43. GrandePre T, Nakamura F, Vartanian T, Strittmatter SM. Identification of the Nogo inhibitor of axon regeneration as a reticulon protein. Nature 2000;403:439–444.
44. von Poser C, Sudhof TC. Synaptotagmin 13: Structure and expression of a novel synaptotagmin. Eur J Cell Biol 2001;80:41–47.
45. Legouis R, Hardelin JP, Levilliers J, Claverie JM, Compain S, Wunderle V, Millasseau P, Le Paslier D, Cohen D, Caterina D, et al. The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. Cell 1991;67: 423–435.
46. Franco B, Guioli S, Pragliola A, Incerti B, Bardoni B, Tonlorenzi R, Carrozzo R, Maestrini E, Pieretti M, Taillon-Miller P, et al. A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. Nature 1991;353:529–536.
47. Kidd T, Brose K, Mitchell KJ, Fetter RD, Tessier-Lavigne M, Goodman CS, Tear G. Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell 1998;92:205–215.
48. Tse W, Zhu W, Chen HS, Cohen A. A novel gene, AF1q, fused to MLL in t(1;11)(q21;q23), is specifically expressed in leukemic and immature hematopoietic cells. Blood 1995;85: 650–656.
49. Tewari M, Yu M, Ross B, Dean C, Giordano A, Rubin R. AAC-11, a novel cDNA that inhibits apoptosis after growth factor withdrawal. Cancer Res 1997;57:4063–4069.
50. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev 1999;13:1924–1935.
51. Yarden RI, Brody LC. BRCA1 interacts with components of the histone deacetylase core complex. Proc Natl Acad Sci USA 1999;96:4983–4988.
52. Bang Y-J, Pirnia F, Fang W-G. Terminal neuroendocrine differentiation of human prostate carcinoma cells in response to increased intracellular cyclic AMP. Proc Natl Acad Sci USA 1994;91:5330–5334.
53. Cox ME, Deeble PD, Lakhani S, Parsons SJ. Acquisition of neuroendocrine characteristics by prostate tumor cells is reversible: Implications for prostate cancer progression. 1999;59: 3821–3830.
54. Burchardt T, Burchardt M, Chen MW, Cao Y, de la Taille A, Shabsigh A, Hayek O, Dorai T, Buttyan R. Transdifferentiation of prostate cancer cells to a neuroendocrine cell phenotype in vitro and in vivo. J Urol 1999;162:1800–1805.
55. Stein CA. Mechanisms of action of taxanes in prostate cancer. Seminars in Oncology 1999;26(Suppl 17):3–7.

218–220 Nucleic Acids Research, 2002, Vol. 30, No. 1

The human (PEDB) and mouse (mPEDB) Prostate Expression Databases

Peter S. Nelson[1,2,*], Colin Pritchard[1], Denise Abbott[1] and Nigel Clegg[1]

[1]Division of Human Biology and [2]Division of Clinical Research, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA

Received September 17, 2001; Accepted September 18, 2001

ABSTRACT

The Prostate Expression Databases (PEDB and mPEDB) are online resources designed to allow researchers to access and analyze gene expression information derived from the human and murine prostate, respectively. Human PEDB archives more than 84 000 Expressed Sequence Tags (ESTs) from 38 prostate cDNA libraries in a curated relational database that provides detailed library information including tissue source, library construction methods, sequence diversity and sequence abundance. The differential expression of each EST species can be viewed across all libraries using a Virtual Expression Analysis Tool (VEAT), a graphical user interface written in Java for intra- and inter-library sequence comparisons. Recent enhancements to PEDB include (i) the development of a murine prostate expression database, mPEDB, that complements the human gene expression information in PEDB, (ii) the assembly of a non-redundant sequence set or 'prostate unigene' that represents the diversity of gene expression in the prostate, and (iii) an expanded search tool that supports both text-based and BLAST queries. PEDB and mPEDB are accessible via the World Wide Web at http://www.pedb.org and http://www.mpedb.org.

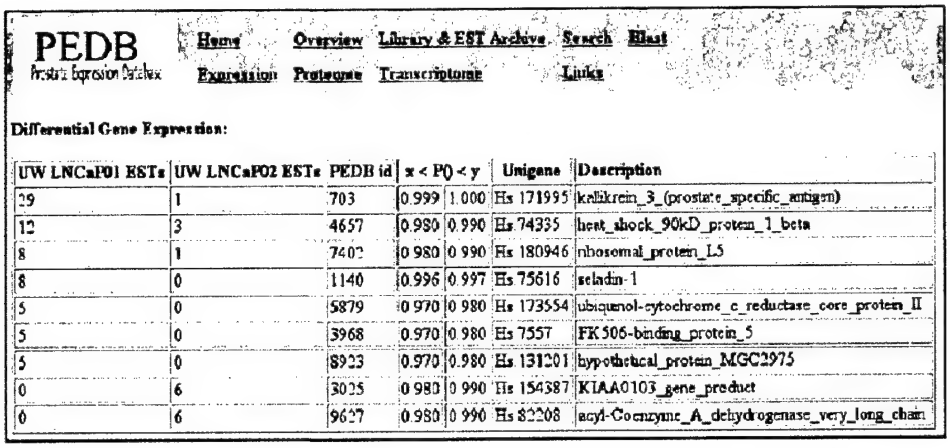
INTRODUCTION

Diseases of the prostate are among the most common pathologies to afflict aging men. Prostate carcinoma is the most frequently diagnosed non-cutaneous malignancy in the US with more than 180 000 new cases estimated for 2001 (1). In order to characterize molecular alterations that accompany prostate disease processes and provide resources for virtual and physical analyses, we have developed the Prostate Expression Database (PEDB) (2). PEDB serves as a centralized collection of gene expression information derived from the human prostate that is organized in a fashion suitable for sequence-based queries, assessment of gene expression diversity, and comparative expression analyses. Expressed Sequence Tags (ESTs) and full-length cDNA sequences derived from 38 human prostate cDNA libraries are archived and represent gene expression profiles reflecting a wide spectrum of normal, benign and malignant prostate disease states. Detailed library information including tissue source, library construction methods, sequence diversity and sequence abundance are maintained in a relational database management system (RDBMS). Prostate ESTs are assembled into distinct species groups using the sequence assembly program Phrap, and annotated with information from the GenBank, dbEST and Unigene public sequence databases.

In recognition of the emerging uses of the mouse as a model system for the study of normal and pathological prostate development, we have developed a database complementary to PEDB that serves to archive and analyze murine prostate gene expression information. The mouse Prostate Expression Database (mPEDB) currently comprises >6000 ESTs from five mouse prostate cDNA libraries constructed from distinct developmental stages and anatomical locations. A detailed description of the database development, data inventory and utilities is available online: www.pedb.org/OVERVIEW/ (Table 1).

Table 1. Table of contents for PEDB overview (http://www.pedb.org/OVERVIEW)

*To whom correspondence should be addressed at: Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, PO Box 19024, Seattle, WA 98109-1024, USA. Tel: +1 206 667 3377; Fax: +1 206 667 2917; Email: pnelson@fhcrc.org



PEDB Prostate Expression Database

Home | Overview | Library & EST Archive | Search | Blast
Expression | Proteome | Transcriptome | Links

Differential Gene Expression:

UW LNCaP01 ESTs	UW LNCaP02 ESTs	PEDB id	x < P0 < y		Unigene	Description
29	1	703	0.999	1.000	Hs.171995	kallikrein_3_(prostate_specific_antigen)
12	3	4657	0.980	0.990	Hs.74335	heat_shock_90kD_protein_1_beta
8	1	7402	0.980	0.990	Hs.180946	ribosomal_protein_L5
8	0	1140	0.996	0.997	Hs.75616	seladin-1
5	0	5879	0.970	0.980	Hs.173554	ubiquinol-cytochrome_c_reductase_core_protein_II
5	0	3968	0.970	0.980	Hs.7557	FK506-binding_protein_5
5	0	8923	0.970	0.980	Hs.131201	hypothetical_protein_MGC2975
0	6	3025	0.980	0.990	Hs.154387	KIAA0103_gene_product
0	6	9627	0.980	0.990	Hs.82208	acyl-Coenzyme_A_dehydrogenase_very_long_chain

Figure 1. Output of differential expression analysis with statistical filtering. The annotated ESTs in two prostate cDNA libraries were compared for relative abundance levels. The output of the analysis provides (i) the number of ESTs in each library corresponding to a specific transcript, (ii) the PEDB identification number, (iii) the statistical probability, P, of differential expression between the two library datasets, (iv) the Unigene database accession number, and (v) a description of the gene based upon GenBank or Unigene annotation.

PEDB DATA AND ANALYSIS TOOLS

PEDB consists of archives of ESTs derived from 38 human prostate cDNA libraries. These ESTs are obtained from public sequence repositories such as GenBank (3), the database of ESTs (dbEST) (4), the Cancer Genome Anatomy Project (CGAP) (5), The Institute for Genome Research (TIGR) or from in-house EST sequencing projects. Sequence processing and curation involves a pipeline of sequence submission, sequence masking, sequence assembly and assembly annotation that now incorporates quality-based assemblies using Phred and Phrap base-calling and sequence assembly algorithms (6,7) (www.pedb.org/OVERVIEW/). Assembled consensus sequences are used for BLAST queries against the Unigene, GenBank and dbEST databases to provide cluster annotation and to further facilitate the assembly process.

The most recent build of PEDB ESTs was assembled starting with 84 832 prostate ESTs. Portions of EST sequences with homology to cloning vectors, *Escherichia coli* genomic DNA and human repetitive DNA sequences were masked. Sequences annotating to the mitochondrial genome were removed and the remaining ESTs with >300 bp of high quality sequence were admitted to the assembly process. A total of 68 426 high-quality ESTs were assembled using Phrap to produce 28 182 clusters. Each cluster was annotated by searching the Unigene, GenBank and dbEST databases using BLASTN. Clusters annotating to the same database sequence were joined to further reduce the number of distinct clusters to 20 187. These annotated assemblies represent the prostate transcriptome: that portion of the genome that is used or expressed in the prostate.

The primary work sites of PEDB involve text-based queries and a BLAST interface for sequence-based searches against PEDB and Unigene datasets. Dynamic gene expression profiles based upon EST assembly and annotation information can be generated using the Virtual Expression Analysis Tool (VEAT). The VEAT provides user-directed inter- and intra-library analysis of transcript abundance, diversity and differential expression. We have recently incorporated a statistical algorithm developed by Audic and Claverie (8) that can determine probabilities of differential transcript abundance levels in datasets comprised of varying numbers of sequences. We have used these tools to identify prostate genes regulated by androgens and genes differentially expressed between adenocarcinoma and small cell carcinoma of the prostate (Fig. 1).

MOUSE PEDB (mPEDB)

The mouse represents a versatile model organism for studying development, genetics, behavior and disease. Several murine models of prostate carcinogenesis have recently been reported (9,10), and the mouse has been used to study the effects of genes hypothesized to be important in the normal and neoplastic development of the human prostate (11). Recognizing the great utility of EST sequences for characterizing organ-specific gene expression, cloning novel genes and developing microarray reagent sets, we have initiated efforts to define the mouse prostate transcriptome by constructing and sequencing mouse prostate cDNA libraries. Interestingly, the extensive list of cDNA libraries provided at the Cancer Genome Anatomy Project web site lists more than 400 murine cDNA libraries, but none are derived from the prostate gland (http://www.ncbi.nlm.nih.gov/ncicgap/).

To date we have made five mouse prostate cDNA libraries, which are derived from microdissected anterior, dorsolateral and ventral prostatic lobes of mature mice, and from the urogenital sinus of E16 embryos. A total of 6145 ESTs have been sequenced, assembled, annotated and loaded into mPEDB in a fashion analogous to that described for processing human prostate sequence in PEDB. Virtual comparisons of transcriptomes derived from these distinct anatomical regions of the prostate suggest that the prostate lobes have specific functional attributes. Library summaries, text- and sequence-based queries, and virtual expression analyses tools are provided.

SUMMARY AND FUTURE DEVELOPMENTS

The human and mouse Prostate Expression Databases serve as centralized archives of gene expression information derived from the human and murine prostate that can be utilized by investigators studying normal and neoplastic prostate development. The assembled human prostate transcriptome currently comprises 20 187 distinct transcripts. Ongoing work involves the characterization of additional cDNA libraries representing specific prostate cell types and early developmental stages, the virtual comparative analyses of human and mouse prostate gene expression, and a database extension for archiving and analyzing cDNA microarray data derived from PEDB and mPEDB sequence resources. PEDB is accessible via the World Wide Web at http://www.pedb.org. mPEDB is accessible at http://www.mpedb.org or through a link from PEDB.

ACKNOWLEDGEMENTS

We thank Alec Mooreman, Stacy Boness and Josh Epstein for sequencing assistance. We thank Burak Eroglu for database administration. This work was supported by the CaPCURE Foundation and grants to P.S.N. from the NCI (K08 CA75173-01A1), NIDDK (RO1 DK59125-02) and the USAMRMC Program in Prostate Cancer Research (DAMD17-98-1-8499).

REFERENCES

1. Greenlee,R.T., Murray,T., Bolden,S. and Wingo,P.A. (2000) Cancer statistics. *CA Cancer J. Clin.*, **50**, 7–33.
2. Hawkins,V., Doll,D., Bumgarner,R., Smith,T., Abajian,C., Hood,L. and Nelson,P.S. (1999) PEDB: the Prostate Expression Database. *Nucleic Acids Res.*, **27**, 204–208.
3. Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J. and Ouellette,B.F.F. (1998) GenBank. *Nucleic Acids Res.*, **26**, 1–7. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 17–20.
4. Boguski,M.S., Lowe,T.M.J. and Tolstoshev,C.M. (1993) dbEST—database for 'expressed sequence tags'. *Nature Genet.*, **4**, 332–333.
5. Schaefer,C., Grouse,L., Buetow,K. and Strausberg,R.L. (2001) A new cancer genome anatomy project web resource for the community. *Cancer J.*, **7**, 52–60.
6. Ewing,B. and Green,P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
7. Gordon,D., Abajian,C., Green,P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.*, **8**, 195–202.
8. Audic,S. and Claverie,J.M. (1997) The significance of digital gene expression profiles. *Genome Res.*, **7**, 986–995.
9. Greenberg,N.M., DeMayo,F., Finegold,M.J., Medina,D., Tilley,W.D., Aspinall,J.O., Cunha,G.R., Donjacour,A.A., Matusik,R.J. and Rosen,J.M. (1995) Prostate cancer in a transgenic mouse. *Proc. Natl Acad. Sci. USA*, **92**, 3439–3443.
10. Di Cristofano,A., De Acetis,M., Koff,A., Cordon-Cardo,C. and Pandolfi,P.P. (2001) Pten and p27KIP1 cooperate in prostate cancer tumor suppression in the mouse. *Nature Genet.*, **27**, 222–224.
11. Bhatia-Gaur,R., Donjacour,A.A., Sciavolino,P.J., Kim,M., Desai,N., Young,P., Norton,C.R., Gridley,T., Cardiff,R.D., Cunha,G.R., Abate-Shen,C. and Shen,M.M. (1999) Roles for Nkx3.1 in prostate development and cancer. *Genes Dev.*, **13**, 966–977.

[CANCER RESEARCH 62, 3920-3924, July 15, 2002]

Advances in Brief

Molecular Effects of the Herbal Compound PC-SPES: Identification of Activity Pathways in Prostate Carcinoma[1]

Michael Bonham, Hugh Arnold, Bruce Montgomery, and Peter S. Nelson[2]

Divisions of Human Biology [M. B., H. A., P. S. N.] and Clinical Research [P. S. N.], Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, and Department of Medicine and Oncology, Veterans Affairs Puget Sound Health Care System, University of Washington, Seattle, Washington 98108 [B. M.]

Abstract

Clinical trials of the herbal preparation PC-SPES have demonstrated substantial responses in patients with advanced prostate cancer. Biochemical assays and clinical observations suggest that the effects of PC-SPES are mediated at least in part through estrogenic activity, although the mechanism(s) remains largely undefined. In this study, we used cDNA microarray analysis to identify gene expression changes in LNCaP prostate carcinoma cells exposed to PC-SPES and estrogenic agents including diethylstilbestrol. PC-SPES altered the expression of 156 genes after 24 h of exposure. Of particular interest, transcripts encoding cell cycle-regulatory proteins, α- and β-tubulins, and the androgen receptor were down-regulated by PC-SPES. A comparison of gene expression profiles resulting from these treatments indicates that PC-SPES exhibits activities distinct from those attributable to diethylstilbestrol and suggests that alterations in specific genes involved in modulating the cell cycle, cell structure, and androgen response may be responsible for PC-SPES-mediated cytotoxicity.

Introduction

Of the many phytotherapeutic compounds advocated for the prevention or treatment of cancer, the herbal preparation PC-SPES is popular among patients with prostate carcinoma as an alternative to conventional forms of therapy. PC-SPES is available as a dietary supplement and is comprised of extracts from eight different herbs: *Scutellaria baicalensis, Glycyrrhiza glabra, Ganoderma lucidum, Isatis indigotica, Panax* pseudo-ginseng, *Dendranthema morifolium* tzvel, *Rabdosia rebescens,* and *Serenoa repens*. Analyses of the individual herbs comprising PC-SPES reveal the presence of numerous bioactive compounds that include phytoestrogens, flavonoids, alkanoids, triterpenes, polysaccharides, and trace elements. Several studies have reported the *in vitro* and *in vivo* efficacy of PC-SPES against prostate carcinoma (1). A Phase II trial of 33 patients with AD[3] prostate cancer and 37 patients with AI prostate cancer assessed the efficacy and toxicity of PC-SPES in a prospective fashion (2). All AD patients showed declines in PSA levels with a median response duration of 57 weeks. Of patients with AI disease, 54% had a PSA decline with a median time to disease progression of 16 weeks. Despite the clinical use of PC-SPES, few active constituents have been identified, and the mechanisms of antineoplastic activity remain to be determined. Biochemical and clinical studies suggest that the effects of PC-SPES are mediated at least in part through estrogenic activity (3), and unpublished reports indicate that the synthetic estrogen DES is present in some preparations of PC-SPES.[4] However, the complexity of the herbal preparation, comprising perhaps hundreds of distinct compounds, implies that other pathways may also be operative. Clinical data raise the possibility that the responses observed with PC-SPES exceed those expected with estrogen alone (2), and studies of the individual herbs comprising PC-SPES report antiproliferative, antimutagenic, and differentiation-inducing activities in multiple tumor types (4–7).

This study was undertaken to determine the molecular mechanism(s) of PC-SPES activity against prostate carcinoma. We used cDNA microarrays to characterize the transcriptional response of LNCaP prostate cancer cells to PC-SPES and compared the gene expression profile with those induced by DES, estradiol, and the synthetic androgen R1881. The transcriptional alterations resulting from these perturbations indicate that PC-SPES exhibits activities distinct from those attributable to DES and suggest that PC-SPES cytotoxicity may be modulated through genes involved in cell cycle control, cell structure, and the AR.

Materials and Methods

Cell Culture and General Methods. DNA manipulations including transformation, plasmid preparation, gel electrophoresis, and probe labeling were performed according to standard procedures (8). Cell lines obtained from the American Type Culture Collection (Manassas, VA) were LNCaP, DU145, and PC3 (each derived from human prostate carcinomas). Cell lines were propagated according to the instructions of the supplier. For experiments determining PC-SPES-mediated temporal gene expression alterations and those comparing PC-SPES with DES, the growth medium was supplemented with 1 or 5 μl/ml PC-SPES (BotanicLab, Brea, CA), 10 μM DES, 30 μM DES (Sigma), or 5 μl/ml ethanol as control. The PC-SPES lots used for these experiments (lot 5431106 and lot 5431164) do not contain detectable levels of DES as determined by independent laboratory analysis. PC-SPES solubilization was achieved by adding 3.2 g (10 tablets) of PC-SPES to 10 ml of ethanol, incubation for 1 h at 37°C, followed by low-speed centrifugation and filtration with a 0.22 μm filter. DES (Sigma) was solubilized in DMSO. For experiments comparing PC-SPES with R1881, DES, and estradiol, LNCaP cells were transferred into RPMI 1640 with 10% CS-FBS (Gemini Biosystems, Woodland, CA) 24 h before treatments. This medium was replaced with fresh CS-FBS media or CS-FBS supplemented with the synthetic androgen R1881 (10 nM; New England Nuclear Life Science Products Inc., Boston, MA), 10 μM 17β-estradiol (Sigma), 10 μM DES, 5 μl/ml PC-SPES, or 5 μl/ml ethanol as control. Total RNA was isolated at specific time points after cell treatments using Trizol (Life Technologies, Inc.) according to the manufacturer's directions. For the microarray experiments, a reference standard RNA was prepared by combining equal quantities of total RNA isolated from LNCaP, DU145, and PC3 cell lines growing at log phase. RNA derived from a single batch of reference standard was used for every microarray hybridization.

Received 8/2/01; accepted 5/23/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by the CaPCURE Foundation and grants from the National Cancer Institute (CA75173-01A1) and Department of Defense (PC991274) to P. S. N. M. J. B. is supported by a Molecular Training Program in Cancer Research Fellowship (T32 CA09437).

[2] To whom requests for reprints should be addressed, at the Division of Human Biology, Fred Hutchinson Cancer Research Center, Mailstop D4-100, 1100 Fairview Avenue N, Seattle, WA 98109-1024. Phone: (206) 667-3377; E-mail: pnelson@fhcrc.org.

[3] The abbreviations used are: AD, androgen-dependent; AI, androgen-independent; DES, diethylstilbestrol; PSA, prostate-specific antigen; CS-FBS, charcoal-stripped fetal bovine serum; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; AR, androgen receptor; CDK, cyclin-dependent kinase.

[4] http://www.psa-rising.com/medicalpike/pespes/.

Microarray Fabrication, Probe Construction, and Hybridization. cDNA microarrays were constructed as we described previously (9). Briefly, a nonredundant set of 3000 distinct prostate-derived cDNA clones was identified from the Prostate Expression DataBase, a public sequence repository of expressed sequence tag data derived from human prostate cDNA libraries (10). Individual clone inserts were amplified by the PCR, purified, and spotted in duplicate onto Type IV glass microscope slides (Amersham) using a GenII robotic spotting tool (Molecular Dynamics, Sunnyvale, CA; Ref. 9).

Fluorescence-labeled probes were made from 30 μg of total RNA in a reaction volume of 20 μl containing 1 μl of anchored oligo(dT) primer (Amersham); 0.05 mM Cy3-dCTP or Cy5-dCTP (Amersham); 0.05 mM dCTP; 0.1 mM each of dGTP, dATP, and dTTP; and 200 units of Superscript II reverse transcriptase (Life Technologies, Inc.). Reactants were incubated at 42°C for 120 min followed by the hydrolysis of RNA and cDNA probe purification by chromatography (Qiagen, Valencia, CA) as described previously (9). Labeled probes were placed onto a microarray slide with a coverslip, hybridized in a humid chamber at 52°C for 16 h, and washed with SSC gradients. Cy3-labeled cDNA from treated cells was directly compared against Cy5-labeled cDNA from the negative control at each time point. Fluorescent dye labeling was reversed, and a replicate experiment was performed for each sample to control for dye effects.

Image Acquisition and Data Analyses. Fluorescence intensities of the immobilized array targets were measured using a GenII slide scanner (Molecular Dynamics). Quantitative data were obtained with the SpotFinder V 2.4 program.[5] Local background hybridization signals were subtracted before comparing spot intensities and determining expression ratios. For each experiment, each cDNA was represented twice on each slide, and the experiments were performed in duplicate, producing 4 data points/cDNA clone/hybridization probe. Intensity ratios for each cDNA clone hybridized with treated and control probes were calculated. Gene expression levels were considered significantly different between the two conditions if all four replicate spot ratios for a given cDNA demonstrated a ratio >1.5 or <-1.5 by at least 1 SD, and the average signal intensity was >800 intensity units. Correlation coefficients between array hybridization data sets were calculated in Excel (Microsoft Corp., Redmond, WA) and expressed as R values. Selected genes were subjected to hierarchical cluster analysis based on an average linkage clustering algorithm using Gene Cluster software (11). Graphical display of clustered genes was generated by Treeview software (11).

Northern Analysis. Ten μg of total RNA were fractionated on 1.2% agarose denaturing gels and transferred to nylon membranes by a capillary method (8). Blots were hybridized with DNA probes labeled with [α-^{32}P]dCTP by random priming using the Rediprime II random primer labeling system (Amersham) according to the manufacturer's protocol. Filters were imaged and quantitated by using a phosphor-capture screen and Imagequant software (Molecular Dynamics).

Cell Proliferation Assay. Ninety-six-well microtiter plates were seeded with 5000 cells/well, and cells were allowed to adhere overnight, followed by the addition of test compounds for 24 or 72 h. Cell proliferation was measured by replacing the culture media with RPMI 1640 containing 1 mg/ml MTT. Isopropanol was added after a 4-h incubation, and cells were incubated overnight at 37°C. The conversion of yellow MTT to a blue formazon dye product was measured with a Micro-Quant spectrophotometer at 570 nm. The amount of formazon dye is a direct indication of the number of metabolically active cells in the culture. Each data point represents the average of four separate experiments containing 8 wells for each experimental condition.

Western Analysis. Thirty μg of protein were loaded into a precast 4–12% gel (Invitrogen), run, and transferred according to the manufacturer's instructions using the X Cell mini cell/blotting module (Invitrogen). Ponceau stain was added to confirm equal loading and transfer. The membranes were blocked overnight at 4°C in 5% milk/PBS. Anti-AR antibody (PharMingen) was added at a 1:1000 dilution for 1 h in 3% BSA/PBS. Horseradish peroxidase-conjugated antimouse IgG antibody was added at a 1:2000 dilution for 30 min. Signals were detected with a chemiluminescence kit (Pierce).

Results

PC-SPES-induced Alterations in Prostate Gene Expression. We performed cDNA microarray analysis to determine alterations in prostate cancer cell gene expression resulting from exposure to PC-SPES. We chose to focus on genes reproducibly exhibiting a $\geq$1.5-fold change in expression level at any time point after treatment. After 8 h, the transcripts of 19 genes increased, and those of 5 genes decreased. After 48 h, the transcripts of 319 genes were altered, with 144 increased and 175 decreased. It was also apparent that the magnitude of induction or repression increased with time for individual genes (Fig. 1). To assist data interpretation, we placed cDNAs encoding characterized genes into distinct functional categories: cell cycle control, metabolism, apoptosis/cell stress, immune modulation, and androgen regulation (Fig. 1). Genes with other functions and cDNAs encoding uncharacterized genes were not grouped. Hierarchical cluster analysis was performed to determine concordant alterations in gene expression over time in each cohort. The complete list of genes and the measured expression alterations at each time point after PC-SPES exposure are available on the World Wide Web.[6]

PC-SPES treatment decreased the expression of several genes encoding cell structural proteins including α- and *β-tubulin*, *dystroglycan*, and *collagen 12* (Fig. 1). Transcripts encoding *filamen*, *α-catenin*, *α-tropomyosin*, *vimentin*, and *α-1 collagen 16* were increased. PC-SPES generally inhibited the expression of genes involved in cell cycle regulation. Transcripts encoding *cyclin A*, *cyclin D*, *cyclin E*, *cdc-20 cdc25B*, *cdc28*, *cdc46*, *CDK2*, *MAD2*, and *cdc6-regulated* protein were decreased. However, the expression of quiescin and the CDK inhibitor *p21* increased. PC-SPES markedly inhibited the expression of all known androgen-regulated genes present on the microarray. Transcripts encoding *PSA*, *TMPRSS2*, *NKX3.1*, *prostase*, and *hK2* were decreased after 24 h of treatment and further diminished at the 48 h time point. PC-SPES up-regulated several genes reported to be associated with apoptosis: *p21*, *clusterin/TRPM2*, *PEA15*, *Gadd 34*, *Id1*, *DAD1*, and *thioredoxin reductase*. The cDNA encoding Bcl-2 was not present in our microarray clone set, thus specific alterations in this apoptosis-regulatory gene were not determined. In support of potential immunomodulatory properties of PC-SPES, altered levels of *thymosin-β-4*, *prothymosin-α*, MHC class I genes, *monocyte-specific enhancer factor*, *interleukin 1*, *IFN-regulatory factor 1* and *2*, and β2 *microglobulin* mRNAs were detected in the prostate cells. We did not examine the effects of PC-SPES on other cell types likely to be effectors of an immune response (*e.g.*, lymphocytes).

To confirm the microarray results, we performed Northern analysis for 17 genes exhibiting gene expression alterations after PC-SPES treatment. For each gene studied, the transcript alterations as measured by Northern were concordant with the array findings (Fig. 2). Selecting a suitable gene to serve as a Northern loading control was difficult because PC-SPES had such a dramatic effect on the overall cellular gene expression profile. For example, *β-actin* was induced 1.6-fold as determined by cDNA array measurements at 48 h and was induced by 2.0-fold on the Northern study (data not shown). Another commonly used housekeeping gene, *G3PDH*, was repressed 1.7-fold by cDNA array measurements and decreased 3-fold by Northern analysis. Therefore, we used methylene blue staining of 28S and 18S ribosomal RNAs as the most reliable control for equivalent loading.

Comparative Analysis of Gene Expression Profiles Reflecting Cellular Exposure to PC-SPES and DES. We determined qualitative and quantitative gene expression profiles reflecting prostate cellular responses to different concentrations of the synthetic estrogen DES and compared these results to expression profiles reflecting

[5] R. Bumgarner, personal communication.

[6] www.pcdb.org/PCSPES/.

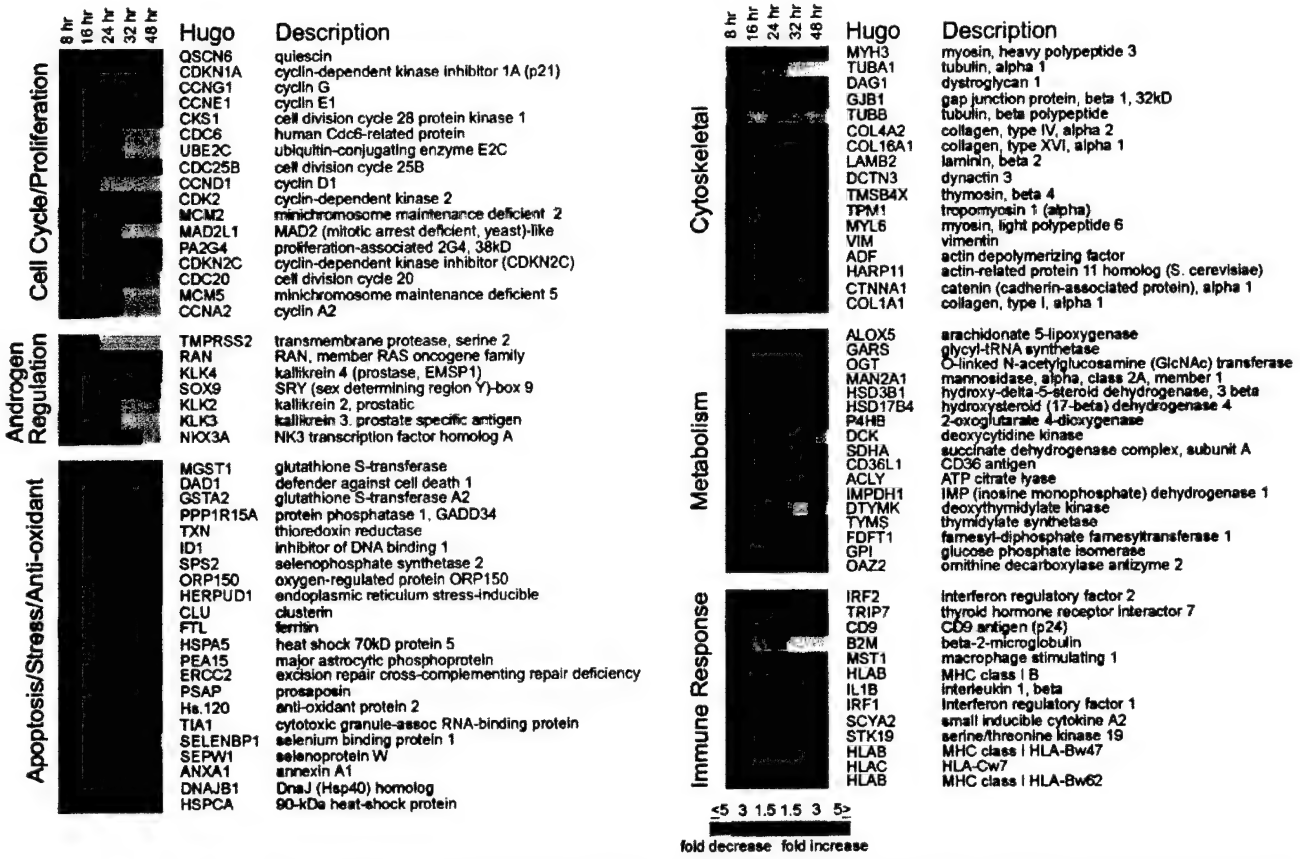


Fig. 1. Temporal alterations in the expression of characterized genes resulting from PC-SPES exposure. Genes are grouped based on known functions and clustered for concordant expression over time. The color scale reflects the experimental fold increase (*red*) or fold decrease (*green*) in transcript abundance relative to the corresponding control experiment.

PC-SPES exposure. DES has been shown to induce apoptosis in LNCaP cells at concentrations between 15 and 30 μM (12). Exposure to PC-SPES also induces apoptosis in LNCaP cells, as well as in AI PC3 and DU145 prostate cancer cell lines (13). To determine cytotoxic equivalence of DES and PC-SPES, LNCaP cells were exposed to different compound concentrations, and cell viability was quantitated using a MTT assay that measures mitochondrial respiratory enzyme activity (14). DES concentrations between 10 and 30 μM (Fig. 3*A*) were equivalent to 3–5 μl/ml PC-SPES in this assay (Fig. 3*B*).

The comparison of global gene expression changes induced by each treatment was performed by plotting the expression change for each gene on the microarray after PC-SPES treatment directly against the corresponding expression change induced by DES (Fig. 3, *C*–*F*). The experimental variability of the microarray assay was demonstrated by hybridizing probes from two independent PC-SPES treatments to two separate sets of microarrays. The coefficient of correlation between these two hybridizations is $r = 0.86$ (Fig. 3*C*). This result demonstrates minimal experimental variation attributable to differences in probe labeling, hybridization, and array construction. Exposure of LNCaP cells to 5 μl/ml PC-SPES for 24 h altered the expression of 156 genes relative to untreated cells (Fig. 3*D*). Treatment with 10 μM DES for 24 h altered the expression of 62 genes. Of these, only six genes (10%) were changed concordantly by PC-SPES. Treatment with 30 μM DES altered the expression of 71 genes, and expression of 12 of these genes (17%) was also changed by PC-SPES. The correlation coefficients between 5 μl/ml PC-SPES and 10 or 30 μM DES are $r = 0.112$ and $r = 0.223$, respectively (Fig. 3, *E* and *F*).

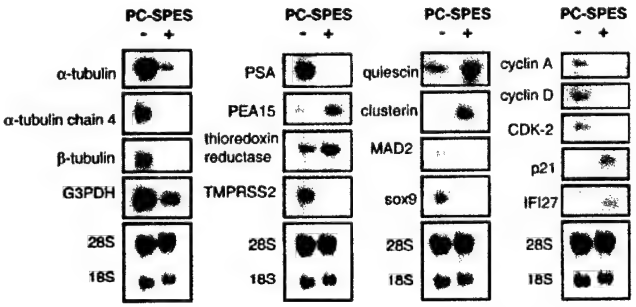


Fig. 2. Northern analysis confirming PC-SPES-mediated gene expression alterations detected by microarray analysis. Equivalent RNA loading is confirmed by methylene blue staining of 28S and 18S RNase. *G3PDH*, glycerol-3-phosphate dehydrogenase; *PEA15*, phosphoprotein enriched in astrocytes 15; *TMPRSS2*, transmembrane protease serine 2; *MAD2*, mitotic arrest-deficient-like 2; *sox9*, SRY box-containing gene 9; *CDK-2*, CDK2; *IFI27*, α-IFN-inducible p27.

In addition to DES, we also compared the PC-SPES gene expression profile with those reflecting cellular responses to the synthetic androgen R1881 and estradiol (results available online[6] as supplemental data). To simulate the environment of prostate cancer in a castrated host, these treatments were performed on LNCaP cells grown in androgen-depleted media. A concentration of 10 nM R1881 altered the expression of 76 genes after 24 h of exposure. The calculated correlation coefficient of $r = 0.009$ between androgen treatment and PC-SPES is indicative of their highly divergent transcriptional effects. In androgen-depleted media, the correlation between DES and PC-SPES gene expression remained low with a coefficient of $r = 0.117$, a value consistent with experiments per-

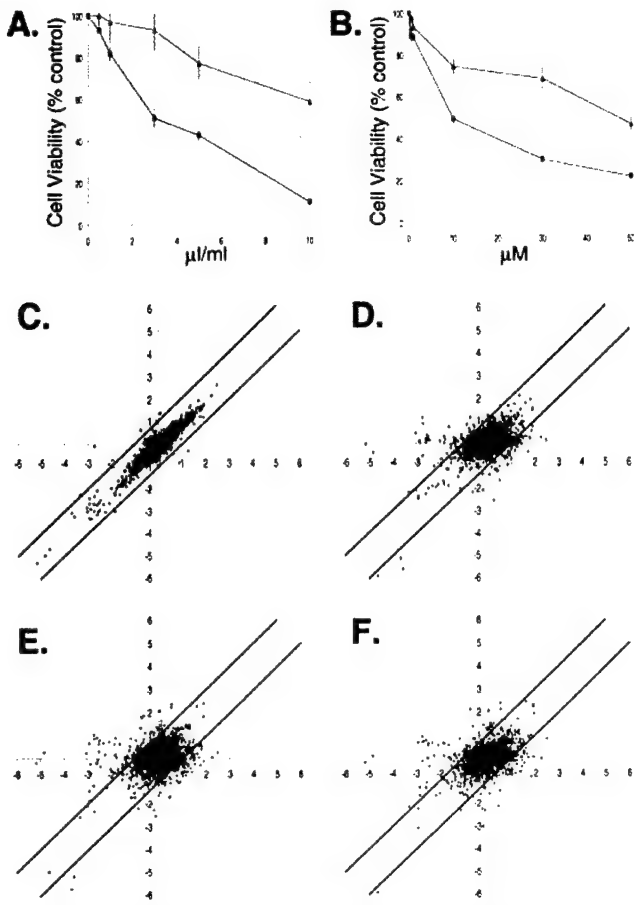


Fig. 3. Comparative analysis of cytotoxicity and gene expression changes resulting from PC-SPES and DES treatment. PC-SPES (*A*)- and DES (*B*)-treated cell viability as measured by MTT assay 24 (▲) and 72 (■) h after treatment. Values are expressed as the percentage of viability of the vehicle control. *C–E*, comparative scatter plots depicting cellular gene expression ratios of PC-SPES treatment against itself (*C*), vehicle control (*D*), 10 μM DES (*E*), or 30 μM DES (*F*). Each point represents the ratio of expression change for a distinct gene plotted for PC-SPES treatment (*X* axis) against the comparison treatment (*Y* axis). Only genes with an average intensity level above background (300 intensity units) are shown.

formed in growth medium containing androgen. Estradiol altered the expression of 49 genes after 24 h when applied at a concentration of 10 μM. A majority of the genes induced by estradiol were also induced by androgen including *PSA, TMPRSS2, hK2,* and *KLK4/prostase*. LNCaP cells are known to express an AR with broad steroid specificity including estrogen-mediated activation (15). When compared with PC-SPES, estradiol exhibited a correlation coefficient of $r = 0.026$.

PC-SPES Regulation of AR Expression. The PC-SPES-mediated transcriptional alteration of several genes known to be androgen regulated prompted additional studies to ascertain whether a common mechanism of control was operative. Northern analysis was performed to determine whether the expression of the AR was changed with PC-SPES treatment. AR transcripts decreased 3–4-fold after 16 h of exposure to PC-SPES, and AR transcripts were undetectable after 48 h of treatment (Fig. 4*A*). The AR message was unchanged over the same time period in the untreated cells. Western blot analysis confirmed that AR protein levels are decreased to undetectable levels 24 and 48 h after treatment of cells with 5 μl/ml PC-SPES (Fig. 4*B*). AR message levels were not significantly reduced by treatment with DES or estradiol, and the addition of androgen did not induce AR transcription in the presence of PC-SPES (Fig. 4*C*). These findings support the microarray data indicating that PC-SPES exhibits activities operating through mechanisms distinct from those attributable to known estrogens.

Discussion

In vitro and *in vivo* studies suggest that multiple biochemical processes are influenced by PC-SPES. A critical metabolic pathway modulating prostate cellular growth involves the interaction of androgenic hormones with the cellular AR. The administration of estrogenic agents such as DES results in castrate levels of serum testosterone through the suppression of the hypothalamic-pituitary-gonadal axis (16). Estrogenic activities of PC-SPES preparations have been documented using *in vitro* assays (3), and patients taking PC-SPES exhibit clinical features consistent with exogenous estrogen administration (2). Thus, a component of PC-SPES efficacy likely results from the suppression of testosterone to castrate levels, an event that occurs in >90% of PC-SPES-treated men with AD disease (2). However, PC-SPES also exhibits activity against AI prostate cancer. In this report, we have shown that gene expression profiles reflective of PC-SPES activity *in vitro* are distinct from profiles of the estrogenic compound DES. Thus, PC-SPES-mediated tumor responses may result both from estrogen-mediated central androgen suppression and direct cytotoxicity via estrogen-independent mechanisms. This conclusion is supported by reports describing PC-SPES activity against AI cells derived from lymphoma and lung carcinoma (17, 18).

The gene expression profiles representing PC-SPES activity indicate several pathways that could contribute to cellular growth inhibition. PC-SPES altered the expression of several genes involved in cell cycle regulation and cell proliferation. Transcripts encoding *CDK2, MAD2*, several orthologues of yeast CDKs, and the G_1 cyclins A, D, and E were significantly reduced. Transcripts encoding *p21*, a protein inhibiting cell cycle progression, were increased by PC-SPES. Taken together, these findings provide further molecular data to support

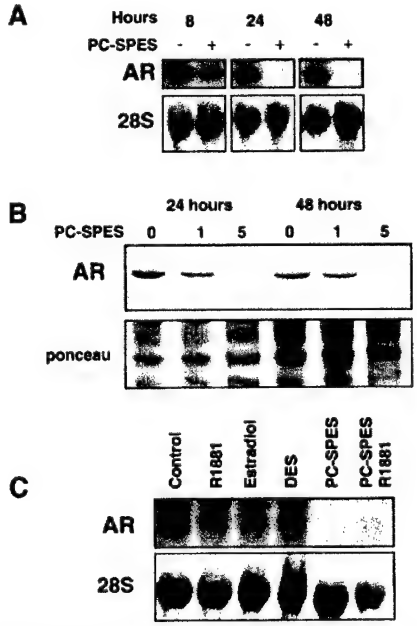


Fig. 4. The effect of PC-SPES on AR expression. *A*, Northern analysis demonstrating down-regulation of AR transcripts in LNCaP cells after 8, 24, and 48 h of exposure to 5 μl/ml PC-SPES. *B*, Western analysis demonstrating down-regulation of AR protein in LNCaP cells after 24 h of treatment with 1 and 5 μl/ml PC-SPES. Ponceau staining is shown as a control for protein loading. *C*, Northern analysis demonstrating AR transcript levels in LNCaP cells after 24 h of treatment with vehicle control, the synthetic androgen R1881, estradiol, DES, PC-SPES, and PC-SPES with R1881.

previous reports describing the antiproliferative effects of PC-SPES including up-regulation of p21 expression and growth arrest at the G_2-M phase of the cell cycle (1). In addition to the observed cell cycle alterations, components of PC-SPES have been shown to initiate an apoptotic response in prostate cancer cells. Licochalcone A, an estrogenic flavonoid extracted from licorice root, has been shown to down-regulate Bcl-2 expression and induce apoptosis in leukemia and breast cancer cell lines (19). Although licorice root is used in the formulation of PC-SPES, it represents only a very minor component,[7] and studies by Kubota *et al.* (1) did not demonstrate alterations of cellular Bcl-2 levels in LNCaP cells treated with PC-SPES. These findings suggest that some mechanisms of PC-SPES cytotoxicity may be cell type dependent.

PC-SPES treatment resulted in the suppression of a large cohort of androgen-regulated genes that included *PSA, hK2, NKX3.1,* and *TMPRSS2*. Several clinical trials have reported a reduction of serum PSA levels in patients taking PC-SPES. Whereas this effect could be mediated through a decline in circulating androgens, we have shown that PC-SPES markedly down-regulates expression of the AR. This finding may account for some of the PC-SPES benefits seen in AI cancers. Several reports have described a cross-talk between the AR and signaling networks such as mitogen-activated protein kinase, and protein kinase A and protein kinase C pathways (20). The reduction of cellular AR by PC-SPES could impair these alternative mechanisms of activating AR-responsive processes. Recent studies of baicalin (21), a flavonoid component of PC-SPES, and of quercitin (4), a flavanoid present in tea and red wine, have shown that each agent can independently down-regulate AR expression. Additional studies of these compounds may serve to characterize new forms of antiandrogen therapy.

In addition to modulating the expression of genes in the AR pathway and those directly involved in cell cycle control, PC-SPES markedly decreased the expression of α- and β-*tubulins*. Tubulin isotypes are structural components of microtubule assemblies that are essential for maintaining cell shape, cell transport, cell motility, and cell division (22). Several chemotherapeutic drugs active against prostate cancer including the taxanes and estramustine function in part through the impairment of microtubule organization and polymerization (23). It is possible that a reduction of cellular tubulins by PC-SPES could provide either a complementary or antagonistic effect toward these and other tubulin-modulating drugs. Additional studies combining PC-SPES with these agents may serve to delineate their optimal use in the clinical setting.

Acknowledgments

We thank Barbara Trask and Steve Plymate for critical reviews of the manuscript. We thank Roger Bumgarner for technical assistance and for array analysis software.

References

1. Kubota, T., Hisatake, J., Hisatake, Y., Said, J. W., Chen, S. S., Holden, S., Taguchi, H., and Koeffler, H. P. PC-SPES: a unique inhibitor of proliferation of prostate cancer cells *in vitro* and *in vivo*. Prostate, *42:* 163–171, 2000.
2. Small E. J., Frohlich, M. W., Bok, R., Shinohara, K., Grossfeld, G., Rozenblat, Z., Kelly, W. K., Corry, M., and Reese, D. M. Prospective trial of the herbal supplement PC-SPES in patients with progressive prostate cancer. J. Clin. Oncol., *18:* 3595-3603, 2000.
3. DiPaola, R. S., Zhang, H., Lambert, G. H., Meeker, R., Licitra, E., Rafi, M. M., Zhu, B. T., Spaulding, H., Goodin, S., Toledano, M. B., Hait, W. N., and Gallo, M. A. Clinical and biologic activity of an estrogenic herbal combination (PC-SPES) in prostate cancer. N. Engl. J. Med., *339:* 785–791, 1998.
4. Xing, N., Chen, Y., Mitchell, S. H., and Young, C. Y. Quercetin inhibits the expression and function of the androgen receptor in LNCaP prostate cancer cells. Carcinogenesis (Lond.), *22:* 409–414, 2001.
5. Matsuzaki, Y., Kurokawa, N., Terai, S., Matsumura, Y., Kobayashi, N., and Okita, K. Cell death induced by baicalein in human hepatocellular carcinoma cell lines. Jpn. J. Cancer Res., *87:* 170–177, 1996.
6. Ghosh, D., Wawrzak, Z., Pletnev, V., Erman, M., Duax, W. L., Pangborn, W., Zhu, D. W., Labrie, F., and Lin, S. X. Molecular mechanism of inhibition of steroid dehydrogenases by licorice-derived steroid analogs in modulation of steroid receptor function. Ann. N. Y. Acad. Sci., *761:* 341–343, 1995.
7. Wang, S. Y., Hsu, M. L., Hsu, H. C., Tzeng, C. H., Lee, S. S., Shiao, M. S., and Ho, C. K. The anti-tumor effect of *Ganoderma lucidum* is mediated by cytokines released from activated macrophages and T lymphocytes. Int. J. Cancer, *70:* 699–705, 1997.
8. Sambrook, J., Fritsch, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.
9. Lin, B., Ferguson, C., White, J. T., Wang, S., Vessella, R., True, L. D., Hood, L., and Nelson, P. S. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res., *59:* 4180–4184, 1999.
10. Hawkins, V., Doll, D., Bumgarner, R., Smith, T., Abajian, C., Hood, L., and Nelson, P. S. PEDB: the Prostate Expression Database. Nucleic Acids Res., *27:* 204–208, 1999.
11. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA, *95:* 14863–14868. 1998.
12. Robertson, C. N., Roberson, K. M., Padilla, G. M., O'Brien, E. T., Cook, J. M., Kim, C. S., and Fine, R. L. Induction of apoptosis by diethylstilbestrol in hormone-insensitive prostate cancer cells. J. Natl. Cancer Inst. (Bethesda), *88:* 908–917, 1996.
13. de la Taille, A., Buttyan, R., Hayek, O., Bagiella, E., Shabsigh, A., Burchardt, M., Burchardt, T., Chopin, D. K., and Katz, A. E. Herbal therapy PC-SPES: *in vitro* effects and evaluation of its efficacy in 69 patients with prostate cancer. J. Urol., *164:* 1229–1234, 2000.
14. Alley, M. C., Scudiero, D. A., Monks, A., Hursey, M. L., Czerwinski, M. J., Fine, D. L., Abbott, B. J., Mayo, J. G., Shoemaker, R. H., and Boyd, M. R. Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. Cancer Res., *48:* 589–601, 1988.
15. Veldscholte, J., Berrevoets, C. A., Ris-Stalpers, C., Kuiper, G. G., Jenster, G., Trapman, J., Brinkmann, A. O., and Mulder, E. The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. J. Steroid Biochem. Mol. Biol., *41:* 665–669, 1992.
16. Kitahara, S., Yoshida, K., Ishizaka, K., Kageyama, Y., Kawakami, S., Tsujii, T., and Oshima, H. Stronger suppression of serum testosterone and FSH levels by a synthetic estrogen than by castration or an LH-RH agonist. Endocr. J., *44:* 527–532, 1997.
17. Hsieh, T. C., Ng, C., Chang, C. C., Chen, S. S., Mittleman, A., and Wu, J. M. Induction of apoptosis and down-regulation of bcl-6 in mutu I cells treated with ethanolic extracts of the Chinese herbal supplement PC-SPES. Int. J. Oncol., *13:* 1199–1202, 1998.
18. Sadava, D., Ahn, J., Zhan, M., Pang, M. L., Ding, J., and Kane, S. E. Effects of four Chinese herbal extracts on drug-sensitive and multidrug-resistant small-cell lung carcinoma cells. Cancer Chemother. Pharmacol., *49:* 261–266, 2002.
19. Rafi, M. M., Rosen, R. T., Vassil, A., Ho, C. T., Zhang, H., Ghai, G., Lambert, G., and DiPaola, R. S. Modulation of bcl-2 and cytotoxicity by licochalcone-A, a novel estrogenic flavonoid. Anticancer Res., *20:* 2653–2658, 2000.
20. Sadar, M. D. Androgen-independent induction of prostate-specific antigen gene expression via cross-talk between the androgen receptor and protein kinase A signal transduction pathways. J. Biol. Chem., *274:* 7777–7783, 1999.
21. Ikezoe, T., Chen, S. S., Heber, D., Taguchi, H., and Koeffler, H. P. Baicalin is a major component of PC-SPES which inhibits the proliferation of human cancer cells via apoptosis and cell cycle arrest. Prostate, *49:* 285–292, 2001.
22. Downing, K. H., and Nogales, E. Tubulin structure: insights into microtubule properties and functions. Curr. Opin. Struct. Biol., *8:* 785–791, 1998.
23. Stearns, M. E., and Tew, K. D. Estramustine binds MAP-2 to inhibit microtubule assembly *in vitro*. J. Cell Sci., *89:* 331–342, 1988.

[7] S. Chen, personal communication.

ELSEVIER

Gene 293 (2002) 149–160

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Isolation and characterization of the murine prostate short-chain dehydrogenase/reductase 1 (Psdr1) gene, a new member of the short-chain steroid dehydrogenase/reductase family

Stacy Moore[a], Colin Pritchard[a], Biaoyang Lin[b], Camari Ferguson[a], Peter S. Nelson[a,c,*]

[a]*Division of Human Biology, Fred Hutchinson Cancer Research Center, Mailstop D4-100, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA*
[b]*The Institute for Systems Biology, Seattle, WA, USA*
[c]*Clinical Research, Fred Hutchinson Cancer Research Center, Mailstop D4-100, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA*

Received 10 December 2001; received in revised form 23 April 2002; accepted 18 May 2002
Received by J.L. Slightom

Abstract

We report the isolation and characterization of a complementary DNA (cDNA) encoding a novel member of the short-chain dehydrogenase/reductase (SDR) gene family that we have designated murine prostate short-chain dehydrogenase/reductase 1 (*Psdr1*). *Psdr1* was cloned as a 3.2 kbp transcript from mouse testis cDNA based on the sequence of the recently described androgen-regulated human *PSDR1* gene (Cancer Res. 61 (2001) 1611). The putative protein encoded by *Psdr1* consists of 316 amino acids with 85% identity to human PSDR1. A search against the BLOCKS database of conserved protein motifs indicates that Psdr1 retains features essential for SDR function. Northern analyses demonstrate that *Psdr1* is highly expressed in the murine testis and liver and exhibits several isoforms. Cloning and sequence analysis of the putative *Psdr1* promoter region identified motifs with homology to the consensus androgen response element and progesterone response element. The *Psdr1* gene was mapped to mouse chromosome 12q31–34, which has synteny with the human *PSDR1* chromosomal location (14q23–24.3). Together, these data describe a new member of the SDR gene family that may be involved in the tissue-specific metabolism of retinoids or steroid hormones. © 2002 Published by Elsevier Science B.V.

Keywords: Short-chain dehydrogenase/reductase; Testis; Steroid; Hydroxysteroid dehydrogenase

1. Introduction

1.1. The short-chain dehydrogenase/reductase gene family

The short-chain dehydrogenase/reductase (SDR) superfamily comprises a large group of functionally diverse proteins expressed in prokaryotes and eukaryotes spanning bacteria to mammals. Although different SDR family members may exhibit amino acid residue identities of only 20–30%, two domains are highly conserved and reflect components of structural and functional significance. Site-directed mutagenesis and crystallographic analyses reveal a common N-terminal motif involved in NAD(H) (Nicotinamide Adenine Dinucleotide) or NADP(H) (Nicotinamide Adenine Dinucleotide Phosphate) co-factor binding, and a Tyr-XXX-Lys sequence involved in the topology of the active site (Jornvall et al., 1995). SDR proteins mediate the metabolism of a wide range of substrates including steroids, flavonoids, retinoids, aldehydes, ketones, sugars, and polycyclic aromatic hydrocarbons, and thus may serve to modulate intercellular and intracellular signaling pathways. Several members of the SDR family catalyze steps in steroid hormone biosynthesis or degradation. These enzymes exhibit distinct tissue expression patterns and activities. Enzymatic modification of steroids has an important role in the regulation of steroid-mediated gene transcription since the balance between active and inactive ligand determines the amount of the signal that can bind and activate nuclear steroid receptors in the target tissue.

Recently, we have isolated and characterized a novel human gene, prostate short-chain dehydrogenase reductase 1 (*PSDR1*) with homology to the SDR family of enzymes (Lin et al., 2001). *PSDR1* expression is primarily localized to the prostate epithelium and is regulated by androgens in the LNCaP prostate cancer cell line (Lin et al., 2001). Of

Abbreviations: SDR, short-chain dehydrogenase/reductase; PSDR1, prostate short-chain dehydrogenase/reductase 1; ARE, androgen response element; PRE, progesterone response element; HSD, hydroxysteroid dehydrogenase; RACE, rapid amplification of cDNA ends; UTR, untranslated region; nt, nucleotides; kbp, kilobase pairs; PCR, polymerase chain reaction

* Corresponding author. Tel.: +1-206-667-3377; fax: +1-206-667-2917.
E-mail address: pnelson@fhcrc.org (P.S. Nelson).

0378-1119/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0378-1119(02)00718-7

relevance for the study of androgen-mediated effects in the prostate gland and other androgen-responsive tissues is the classification of several key enzymes involved in steroid hormone biosynthesis, hydroxysteroid dehydrogenases (HSDs), within the SDR family. This group of HSDs includes 17β-HSD types 1–4 and 6–8 (Peltoketo et al., 1999), 15-hydroxyprostaglandin dehydrogenase (Krook et al., 1990), and 11β-HSD (Oppermann et al., 1997b). 17β-HSD is responsible for the conversion of androstenedione to its more potent form, testosterone (Norman and Litwack, 1997; Zhou and Speiser, 1999; Couture et al., 1993). PSDR1 exhibits significant homology to several of the17β-HSD isoforms (Lin et al., 2001).

Our objective in this study was to isolate the murine ortholog of the human *PSDR1* gene in order to gain additional insights into the evolution and physiology of the SDR gene family, and to further characterize the role of PSDR1 in steroid-responsive tissues. By convention, we have named this new member of the mouse SDR gene family *Psdr1*. In this communication, we report the cloning and characterization of the full-length *Psdr1* complementary DNA (cDNA) and provide information about *Psdr1* expression, chromosomal location, and the evolutionary relationship of *Psdr1* to other SDR family members.

2. Materials and methods

2.1. Cell culture

The TM3 mouse Leydig cell line, TM4 mouse Sertoli cell line, and AML12 mouse hepatocyte cell line were obtained from the American Type Culture Collection (Manassas, VA). TM3 cells were grown in a 1:1 mixture of F-12K medium and Dulbecco's modified Eagle medium (DMEM), supplemented with 4 mM glutamine and adjusted to contain 1.5 g/l sodium bicarbonate, 4.5 g/l glucose, 1 mM sodium pyruvate, 100 U/ml penicillin, 100 μg/ml streptomycin, and 10% fetal bovine serum (FBS). TM4 cells were grown in a 1:1 mixture of modified DMEM and Ham's F12 medium supplemented with sodium bicarbonate, glutamine, glucose, antibiotics, and FBS as above. AML12 cells were grown in a 1:1 mixture of modified DMEM and Ham's F12 medium, supplemented with antibiotics, glutamine, and FBS, as above. All cell culture reagents were purchased from Life Technologies. Cells were harvested at 90% confluence for RNA extraction. Total RNA was isolated using the TRIzol reagent (Life Technologies) according the manufacturer's protocol.

2.2. Psdr1 cDNA isolation

Total RNA was isolated from the testis of sexually mature C57BL/6 mice using the TRIzol reagent. Five-prime rapid amplification of cDNA ends (RACE)-Ready and 3′RACE-Ready cDNA was constructed from mouse testis RNA using the SMART (Switching Mechanism At 5′end of RNA Template) RACE cDNA Amplification Kits (CLONTECH, Palo Alto, CA). Mouse Psdr1 was then amplified using the provided universal primer (CLONTECH) and a primer sequence identical to human PSDR1 at nucleotides 2173–2196 (5′-AGCACACTCCAAACAAGTGATGGG-3′). This human sequence was subsequently shown to have 100% identity with the mouse sequence at nucleotides 3089–3112. The polymerase chain reaction (PCR) product of approximately 3 kbp was subcloned into pCR2.1-TOPO (TOPO refers to Topoisomerase I based cloning) with the TOPO TA cloning kit (Invitrogen) and sequenced. To verify the 5′ end of the cDNA, 5′-RACE was performed using the gene specific primer m6A4-12.5R.137 (5′-GATGAAGGGAAGAGAGAGCAGAAGCAG-3′) and the universal primer (CLONTECH) according to the manufacturer's instructions. Three prime-RACE was carried out similarly using the gene specific primer m6A4.3RN (5′-AAAGCAATGCAGACCAAGGGTGTCAGG-3′). The RACE products were subcloned into pCR2.1-TOPO with the TOPO TA cloning kit (Invitrogen) and sequenced. The Psdr1 coding sequence was confirmed by using the PCR and primer pairs specific for the identified PSDR1 (FLM6A4.1117U, 5′-GAACCGGGGTGTGTCTAGGATCA-3′ and FLM6A4.L, 5′-GTTAAAGATTGGGTCCTGTCAGTC-3′). Amplified PCR products were then subcloned into pCR2.1-TOPO with the TOPO TA cloning kit and subjected to DNA sequencing.

DNA sequencing was performed by the dideoxy chain-termination method using *Taq* dye primer and dye terminator kits (Applied Biosystems). The nucleotide sequences were analyzed with an ABI 377 automated sequencer (Applied Biosystems). Sequence assemblies and analyses were performed using the Sequencher software program (Gene Codes, Corp.).

2.3. Genomic sequencing of mouse Psdr1 intron/exon splice junctions

Mouse genomic DNA (C57BL/6J, Jackson Laboratory) was used as template to amplify introns 1, 4, and 5 by the PCR. Primers were designed near putative splice junctions, as predicted by alignment of mouse *Psdr1* cDNA with human genomic DNA, and contained M13 sequence to facilitate DNA sequencing of splice junctions. The following primers were used: intron 1: 5′, 5′-GTTTTCCCAGTCACGACGAACCGGGGTGTGTCTAGGAT-3′ and 3′, 5′-AGGAAACAGCTATGACCATCCGGGAAGCTGAACATTAGA-3′ (2.6 kb); intron 4: 5′, 5′-GTTTTCCCAGTCACGACGGATGTGCCCCTACTCGAAGA-3′ and 3′, 5′-AGGAAACAGCTATGACCATTTCTAGCAGCAAATGGGTCA-3′ (1.5 kb); and intron 5: 5′, 5′-GTTTTCCCAGTCACGACGCCACAGCAAACTAGCCAACA-3′ and 3′, 5′-AGGAAACAGCTATGACCATCTGTGCCAGGGTGTACAGAG-3′ (2.5 kb). PCR products were purified from agarose gel preparations by spin column purification (Qiagen) and sequenced.

2.4. Promoter cloning by genomic walking

Genomic DNA sequence upstream of the putative *Psdr1* transcriptional start site was obtained using the Genome Walker Mouse DNA library kit (CLONTECH). Briefly, libraries of adapter-ligated genomic DNA fragments were used as a template for PCR reactions with the *Psdr1* gene-specific primers: 5R 137, 5′-GATGAAGGGAAGAGAGA-GCAGAAGCAG-3′; 5RN 113, 5′-CAGGAATCCGAACA-TCTCAGCACCACC-3′; gwR37, 5′-TCGGACAGTTGG-TGTGGCGG-3′ and gwR32, 5′-AGTTGGTGTGGCG-GAAGATG-3′ according to the manufacturer's instructions. PCR products were cloned into the pCR2.1-TOPO vector and sequenced using M13 forward and M13 reverse primers. In total, 665 bp upstream of the 5′-untranslated region (UTR) of PSDR1 were isolated and sequenced. Sequences were examined for promoter and potential transcriptional start sites using a neural network promoter prediction program (Reese, 1998) and for transcription factor binding sites using the Transcription Element Search Software program (Schug and Christian, 1997).

2.5. Chromosomal localization

The T31 mouse radiation hybrid panel (Research Genetics) was used to map the chromosomal localization of PSDR1 with primers m6A4F (5′-AACGGAAAGG-CAGTAATAGACAG-3′) and m6A4R1 (5′-GAGGTTA-TAGATGGTTGTGGTTG-3′). After 35 cycles of amplification, the reaction products were separated on a 1.2% agarose gel. The resulting product pattern was submitted to the Jackson Laboratory Mapping Panel for determination of the probable chromosomal location. A comparative chromosomal analysis with that of human chromosomal markers was performed using the Linkage Map Build tool available at the Jackson Laboratory website (www.informatics.jax.org).

2.6. Northern analysis

Fifteen micrograms of total RNA were fractionated using 1.2% agarose denaturing gels and transferred to nylon membranes by capillary action (Sambrook et al., 1989). Mouse multiple tissue Northern and master blots were obtained from CLONTECH. Mouse Psdr1 DNA probes were generated from mouse testis cDNA by the PCR using the following primer pairs: 5′: 5′-GAATTGACG-CGGTACTCCTC-3′ and 3′: 5′-TCGCTCTTCAGGTTC-AAGGT-3′ to produce a 331 bp product spanning nucleotides 1206–1536 comprising portions of exons 6 and 7 of the mouse Psdr1 cDNA; or 5′: 5′-ACGGCTTTGAGATGCA-CATTG-3′ and 3′: 5′-TCATAATAGAGGAGTACCGCG-3′ to produce a 321 bp product spanning nucleotides 913–1233 comprising portions of exons 4–6 of the mouse Psdr1 cDNA. DNA probes were labeled with [α-^{32}P]dCTP by random priming using the Prime-It RmT Random Primer Labeling Kit (Stratagene) according to the manufacturer's protocol and blots were hybridized using ExpressHyb hybridization solution following the manufacturer's protocol (CLONTECH). Filters were imaged and quantitated by using a phosphor-capture screen and ImageQuant software (Molecular Dynamics).

2.7. Phylogenetic analysis

Protein sequences of the mouse and human PSDR1 along with 12 members of the mouse and human SDR family were used to construct a phylogenetic tree using the computer software package PHYLIP [Phylogeny Inference Package, (Felsenstein, 1993)]. Mouse and human PSDR1 sequences were trimmed at the ambiguous 5′ end of the protein sequences such that alignment(s) began at mouse Psdr1 amino acid 24 (methionine), a 5′ region of homology. Initially, 250 bootstrapped data sets were created using SEQBOOT. A phylogeny estimate, i.e. the maximum parsimony tree, was then calculated for each of these data sets using PROTPARS, a protein sequence maximum parsimony method, with the jumble option specified such that the input order of sequences used five different orderings. The program CONSENSE was then run, resulting in a majority rule consensus tree depicting the outcome of the analyses.

The software program PAUPSearch [a GCG interface to the tree-searching options in PAUP (Phylogenetic Analysis Using Parsimony)] (Swofford, 1998) was also used in phylogenetic analyses of 17 mouse SDR family members, including mouse Psdr1 and human PSDR1. Sequences were first aligned using the GCG program PILEUP and a cladogram was then constructed in PAUP using a maximum parsimony algorithm and heuristic tree search.

Multiple sequence alignments were performed in PILEUP, using progressive pairwise alignments, and in MacVector (Oxford Molecular Ltd.), using a CLUSTAL W formatted alignment. Tree representations of the similarity relationships used to produce sequence alignments were produced using both PILEUP and MacVector. Species names and GenBank accession numbers for SDR members are as follows: *Mus musculus*: cis-retinol/androgen dehydrogenase, BAB03718; trans-retinol/estrogen dehydrogenase, AAF04761; retinal short-chain dehydrogenase/reductase 1, AAH08980; 11-beta corticosteroid short-chain dehydrogenase, P50172; hydroxysteroid 11-beta hydroxysteroid dehydrogenase 2, NP_032315; 17-beta-hydroxysteroid dehydrogenase 1, NP_034605; 17-beta-hydroxysteroid dehydrogenase 2, NP_032315; 17-beta-hydroxysteroid dehydrogenase 3, NP_032317; 17-beta-hydroxysteroid dehydrogenase 4, NP_032318; 17-beta-hydroxysteroid dehydrogenase 7, NP_034606; 17-beta-hydroxysteroid dehydrogenase 8, P50171; WW-domain oxidoreductase, NP_062519; 3-beta hydroxysteroid dehydrogenase-4, NP_032320; 3-beta hydroxysteroid dehydrogenase-1, NP_032319; acetyl-Coenzyme A dehydrogenase, NP_031409; short chain L-3-hydroxyacyl-CoA dehydrogenase, AAK15008; and human: PSDR1,

AAF89632; 17 beta-hydroxysteroid dehydrogenase type 3, NP_0001888; 17 beta-hydroxysteroid dehydrogenase type 7, AAF09266; and corticosteroid 11-beta-dehydrogenase 1, P29608.

2.8. Real-time PCR analysis of Psdr1 expression in castrated and control mice

Four C57BL/6 mice (at 2 months) were castrated under anesthesia. A second group of four C57BL/6 mice were sham-operated as controls. Mice were sacrificed 3 days following castration by using anesthetic followed by cervical dislocation in accordance with IRB (Institutional Review Board)-approved institutional protocols. Prostates (coagulating glands and dorsal lobes) and livers were immediately harvested and snap-frozen in liquid nitrogen. Frozen tissue was ground with a hammer, homogenized in TRIzol, and extracted according to the manufacturer's protocol. Total RNA from either castrated or control mice was pooled for subsequent studies.

cDNA was generated from 1 to 3 μg of total RNA using oligo(dT) and Superscript II reverse transcriptase (Invitrogen), according to the manufacture's suggested protocol. Primers and salts were removed using a Microcon 30 filter (Amicon). Real-time PCR reactions were done in triplicate per experiment using approximately 5 ng of cDNA, 0.3 μM of each primer, and 1× SYBR Green PCR master mix (Applied Biosystems) in a 50 μl reaction volume. cDNA generation and PCR amplifications were repeated three times. Reactions were performed and analyzed using an Applied Biosystems 5700 sequence detector. Following normalization by measuring the expression (cycle threshold value during exponential amplification) of the control ribosomal gene S16 (Foley et al., 1993; Pritchard et al., 2001), *Psdr1* expression levels were determined and are reported relative to *Psdr1* levels in control mice. Results are reported as the average fold-change from three experiments. Error bars represent the error of the mean from the three experiments. Primer sets were tested using serial 10-fold dilutions of template. For the 10-fold dilutions, the difference in threshold cycle number was approximately 3.2, indicating high PCR efficiency. Control reactions with RNA or water as template did not produce significant amplification products. Amplification of a single PCR-product per reaction was monitored by generation of a single dissociation curve. The sequences of primers used in this study were: S16 forward, 5′-AGGAGCGATTTGCTGGTGTGGA-3′; S16 reverse, 5′-GCTACCAGGCCTTTGAGATGGA-3′; mPS-DR1 forward, 5′-ATCCTGTACTTGGTCACGCCA-3′, and mPSDR1 reverse, 5′-CACCCCACTGGACAGCATTT-3′.

3. Results and discussion

3.1. Isolation and characterization of murine Psdr1

cDNA synthesized from mouse testis RNA was used as template to amplify and clone the murine ortholog of the recently described human PSDR1 gene (Lin et al., 2001). RACE was used to obtain the 5′ and 3′ ends of the complete Psdr1 cDNA sequence (Fig. 1). The composite cDNA spans 3283 nucleotides and encodes a putative protein of 316 amino acids with a theoretical molecular weight of approximately 35 kDa. A 5′-UTR extends 503 nucleotides upstream of the predicted start codon, GAGATGT, which has modest alignment with the Kozak translation initiation consensus sequence (RNNatgG), where R is a purine (Kozak, 1987). The open reading frame begins with the ATG and extends 948 nt. This is followed by a stop codon, TAA, and a 3′ UTR of 1832 nt, terminating in a poly(A) tail. Three potential polyadenylation signals were identified at nucleotide positions 2026, 3215, and 3241, the latter aligning with a potential polyadenylation signal in human PSDR1. PCR primers flanking the start codon and within the 3′-UTR were used to amplify the entire Psdr1 coding region from mouse testis cDNA. As expected, a 2.8 kbp band was produced. DNA sequencing confirmed the size and identity of the amplified DNA (data not shown). The full-length Psdr1 sequence has been deposited in GenBank with accession number AY039032.

Sequence alignments of the complete cDNAs of human and mouse PSDR1 demonstrates an overall identity of 37%, and a 78% identity over the coding regions. Mouse Psdr1 is approximately 700 bp larger than the human PSDR1 transcript. Based on sequence alignment, this is due to more extensive 5′ and 3′UTR sequence. A comparison of the putative human and mouse PSDR1 translations shows that the proteins are 85% identical. A search against the BLOCKS database [www.blocks.fhcrc.org, (Pietrokovski et al., 1996)] indicates that PSDR1 is a member of the SDR family. A multiple sequence alignment of mouse Psdr1 protein sequence, human PSDR1 protein sequence, and isozymes of mouse 17β-hydroxysteroid dehydrogenase is shown in Fig. 2.

Two motifs are highly conserved in the SDR family and these are highlighted in Fig. 2 with asterisks. The first is a GlyXXXGlyXGly segment, which is characteristic of the coenzyme NAD(H) or NADP(H) binding fold in dehydrogenases (Jornvall et al., 1995). The second motif, TyrXXXLys, is thought to be involved in enzyme catalytic activity, a hypothesis supported by mutagenesis experiments (Jornvall et al., 1995; Cols et al., 1993; Chen et al., 1991; Oppermann et al., 1997a). Like the human PSDR1 protein, the mouse Psdr1 also contains these conserved motifs. Overall, SDR family members exhibit residue identities of only 20–30%, suggesting early duplicatory origins and extensive divergence (Jornvall et al., 1995). Searches against PROSITE patterns database [www.isrec.isb-sib.ch/software/PSTSCAN_form.html, (ISREC Bioinformatics Groups et al., 1999)] revealed that the mouse PSDR1 contains an Asn-glycosylation site at amino acid position 171, two protein kinase C (PKC) phosphorylation sites (amino acids 54 and 103), two casein kinase II phosphor-

```
   1 atagggcaaaggtcatctctgaggaaggacatcttccgccacaccaactgtccgactaacggctcctgccctttgtaattaacctcagaatactttccag 100

 101 acgctgtctcccagacaaaatgtaatttcctaataatgattggggaaaattgaaccaaacactggacttgggctccaggagatagcaaacccgcgagact 200

 201 tccaaagagtcaattagtggccagaggagtggtagaagggctggctcaacgatcccacaggcctaggccctggaccactgaagaccattgctccacggtt 300

 301 gtgttattcaacttttcatagtgcttggaagtatactctgaatgtatcaccgccaggcggccgccaaatccaggactacatttcccagagggctttgctc 400

 401 cctggcggttgcagttgaaccggggtgtgtctaggatcagaagccataagtcccgctgtcctctttagagcatcaacttcaactgcgcttcggtggtgct 500

 501 gag ATG TTC GGA TTC CTG CTT CTG CTC TCT CTT CCC TTC ATC CTG TAC TTG GTC ACG CCA AAA ATC AGG AAA ATG  575
   1     M   F   G   F   L   L   L   L   S   L   P   F   I   L   Y   L   V   T   P   K   I   R   K   M   24

 576 CTG TCC AGT GGG GTG TGC ACA TCT AAT GTT CAG CTT CCC GGG AAG GTA GCC ATA GTC ACT GGT GCT AAC ACA GGC  650
  25 L   S   S   G   V   C   T   S   N   V   Q   L   P   G   K   V   A   I   V   T   G   A   N   T   G   49

 651 ATT GGG AAG GAG ACA GCT AAA GAT CTG GCC CAA AGA GGA GCC CGT GTG TAT TTA GCT TGC CGG GAT GTG GAC AAG  725
  50 I   G   K   E   T   A   K   D   L   A   Q   R   G   A   R   V   Y   L   A   C   R   D   V   D   K   74

 726 GGG GAA CTG GCG GCT CGT GAG ATC CAA GCC GTC ACA GGG AAC AGT CAG GTC TTC GTA CGG AAA CTG GAC CTA GCT  800
  75 G   E   L   A   A   R   E   I   Q   A   V   T   G   N   S   Q   V   F   V   R   K   L   D   L   A   99

 801 GAT ACC AAG TCT ATT CGA GCC TTT GCC AAA GAC TTC TTA GCT GAG GAA AAG CAT CTG CAC CTT CTC ATC AAC AAT  875
 100 D   T   K   S   I   R   A   F   A   K   D   F   L   A   E   E   K   H   L   H   L   L   I   N   N  124

 876 GCG GGC GTG ATG ATG TGC CCC TAC TCG AAG ACT GCA GAC GGC TTT GAG ATG CAC ATT GGC GTC AAC CAC CTG GGT  950
 125 A   G   V   M   M   C   P   Y   S   K   T   A   D   G   F   E   M   H   I   G   V   N   H   L   G  149

 951 CAC TTC CTC CTG ACC CAT TTG CTG CTA GAA AAG CTG AAG GAG TCG GCC CCA TCA AGG ATA GTC AAC TTG TCT TCT 1025
 150 H   F   L   L   T   H   L   L   L   E   K   L   K   E   S   A   P   S   R   I   V   N   L   S   S  174

1026 TTG GGA CAC CAT CTG GGC AGG ATC CAC TTC CAT AAC CTG CAG GGG GAG AAG TTC TAC AGT GCG GGT CTC GCG TAC 1100
 175 L   G   H   H   L   G   R   I   H   F   H   N   L   Q   G   E   K   F   Y   S   A   G   L   A   Y  199

1101 TGC CAC AGC AAA CTA GCC AAC ATT CTC TTC ACT AAG GAG CTG GCC AAG AGG CTG AAA GGT TCT GGA GTG ACA ACA 1175
 200 C   H   S   K   L   A   N   I   L   F   T   K   E   L   A   K   R   L   K   G   S   G   V   T   T  224

1176 TAC TCT GTA CAC CCT GGC ACA GTC CAT TCT GAA TTG ACG CGG TAC TCC TCT ATT ATG AGA TGG CTT TGG CAA CTT 1250
 225 Y   S   V   H   P   G   T   V   H   S   E   L   T   R   Y   S   S   I   M   R   W   L   W   Q   L  249

1251 TTC TTT GTT TTC ATC AAG ACC CCT CAA GAG GGA GCT CAG ACG AGC CTG TAC TGT GCC CTG ACA GAA GGT CTC GAG 1325
 250 F   F   V   F   I   K   T   P   Q   E   G   A   Q   T   S   L   Y   C   A   L   T   E   G   L   E  274

1326 AGC CTA AGT GGC AGT CAT TTC AGT GAT TGC CAG TTG GCA TGG GTC TCT TAC CAA GGT CGC AAT GAG ATA ATA GCC 1400
 275 S   L   S   G   S   H   F   S   D   C   Q   L   A   W   V   S   Y   Q   G   R   N   E   I   I   A  299

1401 AGG CGG CTG TGG GAT GTC AGC TGT GAC CTG CTG GGC CTC CCA GTG GAT TGG TAA gtggtggtttggactcaaaagaagattg 1482
 300 R   R   L   W   D   V   S   C   D   L   L   G   L   P   V   D   W   *                             317

1483 gaggagatgatgattatccttcaaagtggccaaaaccttgaacctgaagagcgaagaacttcaagcctttcctgcttggcatccagttaaatcctagtac 1582

1583 actgccaggttcctcgaaacccctgagtttgtattgacttattttgttcctgctcctgccagcgtttctagtggtatcactaagacagaggaccttcat 1682

1683 gtgacctacacagctcctatttttcctttggaaactccccctaaccaggagcatgaaagccctatgaagattggtacatgacaatgtggattagatggagc 1782

1783 tcctccgagccgaccatcctctccttgaagaactccattgtaacctcagctgaacctcaggagtcccggagcctggctcaagggcaggcagggctttgtg 1882

1883 gtcccaacatctggatacttaagaaaagtttttactaaatctggcagtctgaagctttggcttctgagacacttttctatgtccagaccactgaagagct 1982

1983 tccttctctaagaaatttgtgggatttgaagggcagagataaaaataaagccaaactgtcattttctcccgctgtttctcagcctagtaaggaggtaatg 2082

2083 aaaggaacggtatgagacgataggtcactggggtcagttactgaccccagacccacagaggcctctttttgtgccttagtttctcttaggaaactagaa 2182

2183 aaggaaaggccttctcattgtcattgtgtcgattttgtgtgtcttcttgtgcattcatgcagttgaatgtgtttccacaaaattcacaatagccacaaag 2282

2283 gaaaagttcttaatcagactgtagtgtagacaaattaccagagatgggacaaatcactgtggagtctttgagtacagttgcaaagcaagggagacactat 2382

2383 ctgtgttgtaatcggtcactgagaagcacagatgccgcagggaagaagtttaaaagaaaaatcctaggaggggacacatgataatgaaggggtcagctga 2482

2483 ggatctaaaaccataacggaaaggcagtaatagacagctacagaaaggcacactgtcttaatcaccagtcagtcagagtgataggaaaatgtccaaagca 2582

2583 caagagtgaaatatttctctattcagaaatgttatcctaaggattggtaattatagtcagttcataatcttttagagcattttcttacattagcttaaca 2682

2683 agatgtccatttcaggaatgtgtatggagagatgggatggcttagtaatgcctgctgttccagaggaccctggttcaattactaggacccacttggcaac 2782

2783 cacaaccatctataacctcagttatagatgatctagaggatcagacagcctcatctacctccacaggccctgcaggaatgtggtgcatagatatatatgc 2882

2883 aggcacacatacaaataatttttgtaaaaaaaaagaaatcaattttgacttttgggggtatgctctcctgaaatgttgaggcccctgggtttgatcccagca 2982

2983 ctgcaaaagaaagtagagatagcctcaatttactacgattctgctttatttggagagcttttatcaaaagcaatgcagaccaagggtgtcaggaatccct 3082

3083 catgttcccatcacttgtttggagtgtgctattctaagggagtttgccatttcctcgggctgacaattatattttaagcttgaatatgtaagactgacag 3182

3183 gacccaatctttaacttctaatatttgtaaattaataaatttatgggtttgttctgaccaataaacgtattttaaaatgaaaaaaaaaaaaaaaaaaaaa 3282

3283 a                                                                                                    3283
```

Fig. 1. Nucleotide and predicted amino acid sequence of mouse *Psdr1*. The potential initiation methionine codon and the translation stop codon are bold. Potential polyadenylation signals are underlined.

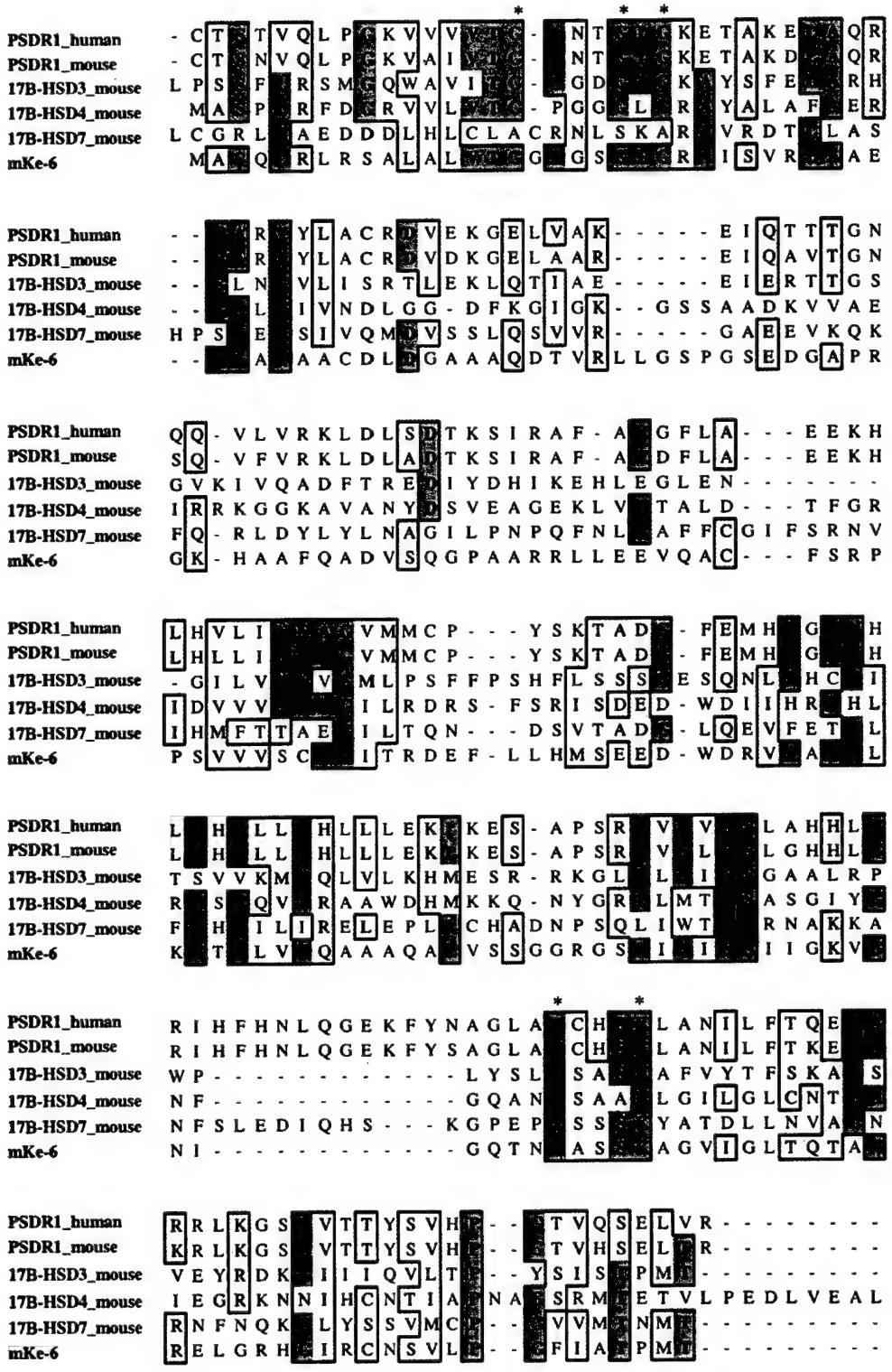


Fig. 2. Multiple sequence alignments of mouse Psdr1 with human PSDR1 and 17β-HSD family members of the SDR super family. The alignment was performed with the CLUSTAL W program using MacVector 6.5 software. BLOSUM series matrix was used with an open-gap penalty score of 10 and an extend-gap penalty score of 0.05. *Boxed* and *dark-shaded*, identical residues; *boxed* and *light-shaded*, similar residues. *, two conserved segments of the SDR family, GlyXXXGlyXGly (coenzyme binding site) and TyrXXXLys (involved in catalytic activity). The GenBank accession numbers for the members aligned here are: human PSDR1, AAF89632; mouse 17β-HSD 3, NP_032317; mouse 17β-HSD 4, NP_032318; mouse 17β-HSD 7, NP_034606; mouse 17β-HSD 8 (m*Ke-6*), P50171. Only regions containing the conserved motifs are shown here.

Table 1
Exon locations and sizes from mouse *Psdr1* as determined by mouse genomic DNA sequencing and sequence alignment with public mouse genomic DNA sequence

Exon	Human *PSDR1* exon size	Mouse *Psdr1* exon size	5′ splice donor	3′ splice acceptor
1	> 114	568	TCAG\gtctg[a]	
2	120	119	AGAG\gcaag	tctag/GAAA[b]
3	159	156	GCTG\gtaag	gacag/GAGCC
4	104	105	CTGG\gtaag[b]	tgcag/AGGAA
5	211	210	AAAG\gtgag[a]	aacag/GTCAC
6	191	191	TCAG\gtatg	cccag/GTTCT[b]
7	1621	1934		tccag/TGATT

[a] Determined by genomic DNA sequencing, sequence not available in the public domain.
[b] Confirmed by genomic DNA sequencing.

ylation sites (amino acids 54 and 256), and seven N-myristoylation sites. The Asn-glycosylation site, PKC phosphorylation sites, and several N-myristoylation sites are shared with the human PSDR1 sequence.

3.2. *PSDR1* genomic organization and promoter sequence analysis

Intron/exon junctions of the mouse *Psdr1* gene were determined by sequence alignment of the mouse *Psdr1* cDNA with mouse genomic DNA (www.genome.ucsc.edu) (Table 1). Junctions were also determined by DNA sequencing as described in Section 2, when genomic sequence was not publicly available. Comparisons of mouse *Psdr1* and human *PSDR1* genomic sequences suggests similar gene structure and exon sizes (Table 1). The *PSDR1* gene compromises seven exons and six introns. Notable differences between the two putative orthologs are the extended lengths of exons 1 and 7 of the mouse *Psdr1* relative to human *PSDR1*. These exons comprise the 5′ and 3′ UTRs, respectively.

Sequences 5′ to the putative *Psdr1* transcriptional start site were cloned and examined for potential transcription factor binding sites using the TESS (Transcription Element Search Software, www.cbil.upenn.edu/tess/index.html) program and MacVector 6.0 software (Fig. 3). Two sequences with greater than 70% homology to the consensus androgen response element [ARE; 5′-GGA/TACAnnnTGTTCT-3′, (Roche et al., 1992)] one sequence with greater than 65% homology to the consensus progesterone response element [PRE, (Lieberman et al., 1993)], as well as an interleukin-6 response element binding protein site, TTCCCAGAA (Hocke et al., 1992), were identified within 660 bp 5′ of the putative transcription initiation site. The putative promoter region of human *PSDR1* also contains potential IL-6, ARE, and PRE motifs, suggesting that the regulation of *PSDR1* expression is likely to be similar in the mouse and human.

3.3. Phylogenetic analysis

Protein sequence alignments using the programs PILEUP and MacVector 6.5 show that mouse Psdr1 is most similar to

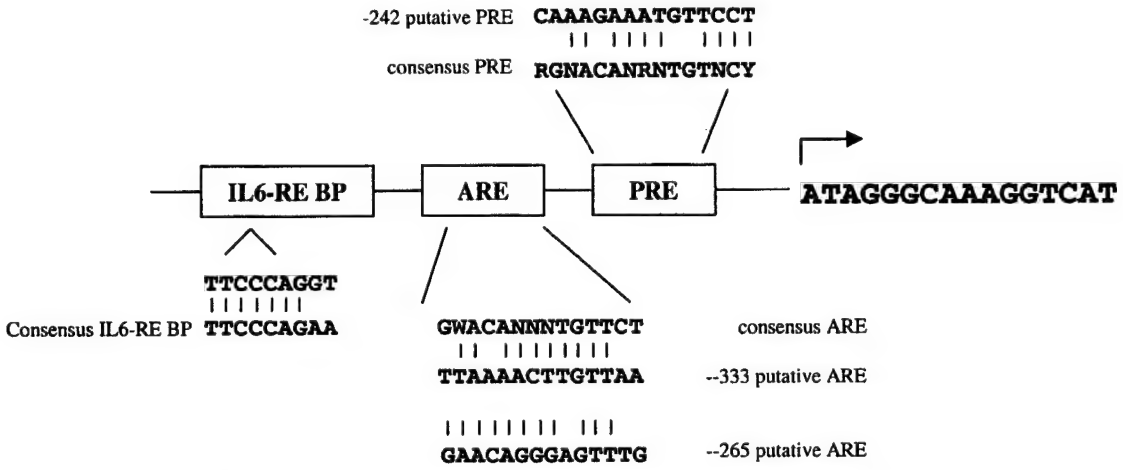


Fig. 3. Schematic drawing demonstrating the putative sequence motifs of the mouse *Psdr1* promoter. Arrow, the predicted transcription initiation site, based on 5′RACE products. Putative ARE and PRE sequences, and an interleukin-6 response element binding protein (IL6-RE BP) site at −462 are also shown. R, purine; Y, pyrimidine; W, A or T; N, any nucleotide.

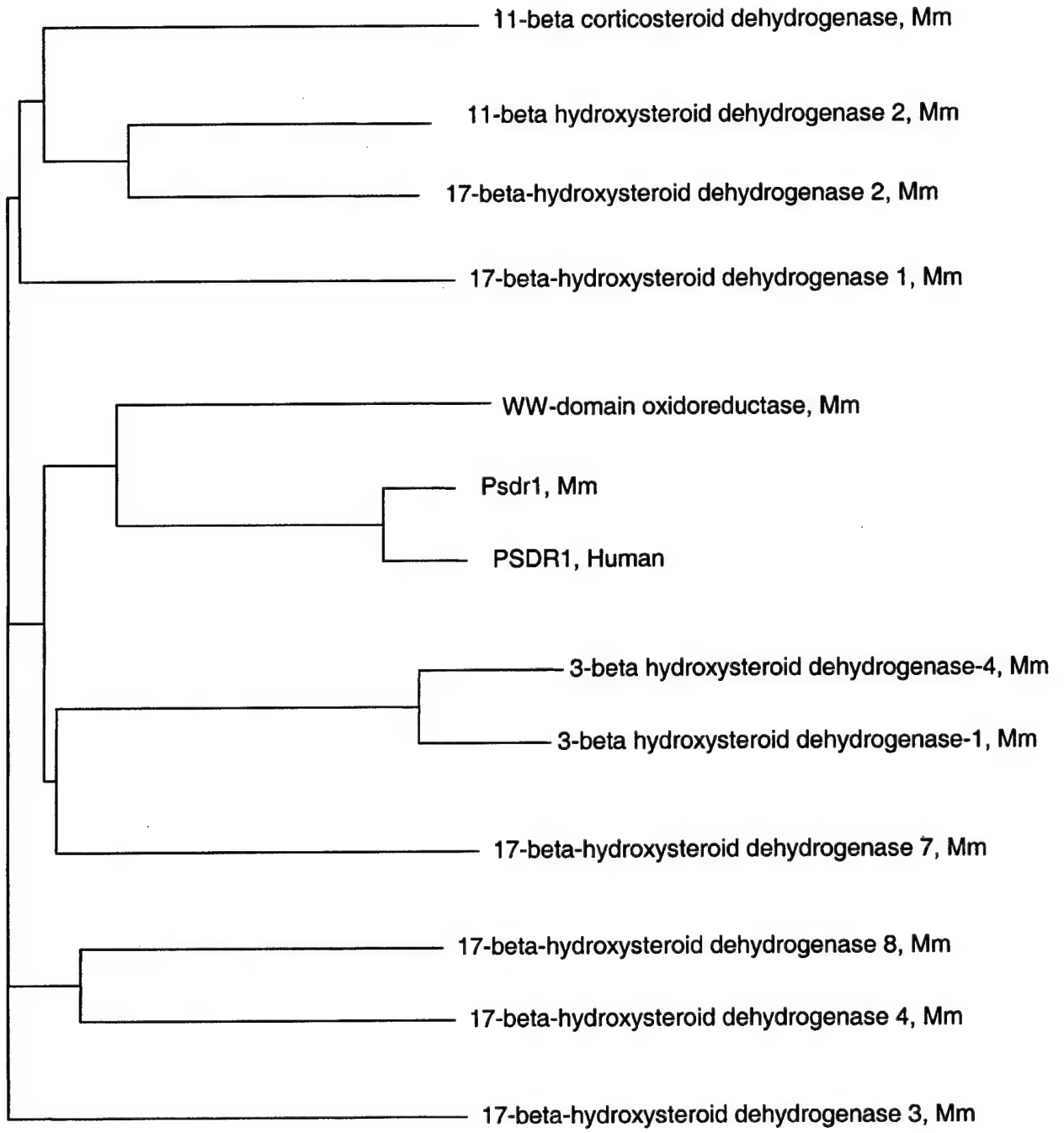


Fig. 4. Dendrogram demonstrating the similarity relationships among thirteen mouse SDR proteins, including, mouse Psdr1, and human PSDR1. The tree was constructed following sequence alignment using the CLUSTAL W program in MacVector 6.5. Mm, *M. musculus*.

human PSDR1 and that these putative orthologs are similar to the mouse WW-domain oxidoreductase, mouse 3β-HSD 1 and 4 and mouse 17β-HSD 7 (Fig. 4). According to bootstrap analyses, using the programs PHYLIP and PaupSearch, of the sequences analyzed, mouse Psdr1 and human PSDR1 cluster together, suggesting that human PSDR1 and the described mouse Psdr1 are orthologs. Human and mouse PSDR1 clustered most closely with mouse 17β-HSD 7 (data not shown) an enzyme involved in ovarian estradiol biosynthesis in luteinized cells (Nokelainen et al., 1996). Also clustering closely with PSDR1 were enzymes involved in retinoid metabolism, including mouse cis-retinol/androgen dehydrogenase type 2 (CRAD2). Thus, it is possible that PSDR1 could be functioning in steroid and/or retinoid metabolism. Spatial and temporal affects, as well as substrate availability likely govern the function of PSDR1.

3.4. Mapping of Psdr1 to chromosome 12

The mouse T31 radiation hybrid panel was used to determine the chromosomal location of Psdr1 using the gene-specific primers, 6A4F and 6A4R1. Analysis of the typing results indicated that Psdr1 had a highest anchor LOD of 11.5 to D12Mit4 between the two markers, D12Mit92 (mapped to 12q31) and D12Mit4 (mapped to 12q34). Thus, Psdr1 is mapped to 12q31–34. Notably, this region has synteny with human chromosome 14q21–24, a region encompassing the mapped position of human PSDR1 (Lin et al., 2001).

Numerous studies support associations between molecular variations in SDR family members and the development and progression of human diseases. For example, germline mutations in 17β-HSD 3 and 4 result in the male pseudohermaphroditism (Geissler et al., 1994) and the fatal form of Zellweger syndrome (de Launoit and Adamski, 1999; Peltoketo et al., 1999), respectively. Abnormal regulation of 17β-HSD8 (*Ke 6*), an alternatively spliced gene member of the SDR family, has been linked to recessive polycystic kidney disease in mice (Aziz et al., 1993). Allelic variants in the 3β-HSD 2 gene, encoding one of two enzymes that initiates the inactivation of dihydrotestosterone, have been identified and are currently under assessment for a role in racial/ethnic differences in prostate carcinogenesis (Devgan et al., 1997). Database and literature searches did not identify diseases or syndromes with linkage to the chromosomal mapping locations of human or mouse PSDR1. However, PSDR1 remains a candidate for evaluation in diseases where hormone metabolism may be a contributing factor.

3.5. Tissue expression profile of Psdr1

The distribution of *Psdr1* transcripts in normal mouse tissues was examined by Northern analysis and messenger RNA (mRNA) dot blot (Figs. 5A,B). *Psdr1* was expressed predominantly in the testis with a message size of approximately 3.5 kbp, similar to the expected size determined by cDNA sequencing. Three additional bands of approximately 2.9, 2.1, and 1.8 kbp, were also observed. The 2.9 and 1.8 kbp transcripts were also expressed in mouse liver. These bands may be due to alternative splicing, use of different *Psdr1* transcription start sites and polyadenylation signals, or cross-hybridization to transcripts with homology to *Psdr1*. Two different, but overlapping, probes corresponding to mouse *Psdr1* exons 4–6 and 6–7 produced identical expression patterns indicating that this region of *Psdr1* (913–1536 bp of *Psdr1* cDNA) is homologous to several mRNA transcripts. The 2.1 kb band likely represents a *Psdr1* transcript utilizing a polyadenylation signal located in exon 7 at nucleotide position 2026. In addition, searches of the database of expressed sequence tags (ESTs) identified several putative *Psdr1* isoforms including ESTs with truncated 5′ UTRs (accession number BG083594), splicing of exon 2 (accession number BI558007) and alternative 3′ UTRs (accession number AW048696). Combinations of these isoforms could account for the *Psdr1* transcript size variations identified by Northern analysis, and indicates a complex transcript processing program that is tissue-type and cell-type dependent.

During the preparation of this manuscript, two sequences with identity to the Psdr1 sequence were deposited into GenBank. One sequence is designated SCALD for short-chain aldehyde dehydrogenase (AF474027) and the other is designated Mdt1 for cell line MC/9.IL4 derived transcript 1, and represents a full-length cDNA cloned from a murine mammary tumor (BC018261). Studies describing these genes have yet to be published. The encoded proteins are identical in both sequence and length to PSDR1 and would not be expected to exhibit different message sizes. A previously deposited sequence published only in GenBank, ube-1a (AB030503), is also identical to nucleotides 144–2048 of the *Psdr1* cDNA. As with SCALD and Mdt1, alternatively spliced forms of ube-1a are not described. Database searches with the *Psdr1* cDNA sequence did not identify additional genes or transcripts that would be expected to cross-hybridize with *Psdr1*. However, searches of the human genome assembly (http://genome.ucsc.edu/) have identified two additional human PSDR1 family members (data not shown). Completion of the mouse genome may identify additional murine *Psdr1* homologs orthologous to these human genes that could account for the additional transcript sizes seen by Northern analysis.

The *Psdr1* tissue expression profile was confirmed using an RNA Master dot blot comprised of mRNAs from 22 different mouse tissues (Fig. 5B). *Psdr1* expression was predominantly expressed in testis with a lower but significant level of expression in liver, smooth muscle, thymus, submaxillary gland, and epdidiymis. *Psdr1* expression was 2-fold higher in testis relative to any other mouse tissue examined. *Psdr1* expression was also detected in days 15 and 17 mouse embryos. In contrast, human *PSDR1* has the greatest level of expression in the prostate, followed by lower expression levels in the liver, testis, kidney and pancreas. Prostate specific membrane antigen, prostate stem cell antigen, and TMPRSS2 are other examples of genes expressed in the human prostate that exhibit different tissue expression patterns in the mouse (Bacich et al., 2001; Ross et al., 2001; Jacquinet et al., 2000). Despite common characteristics such as secretory functions and hormonal regulation, these gene expression variations may reflect the significant anatomical differences between human and mouse prostate.

As the mouse testis and liver expressed higher *Psdr1* levels relative to other tissues studied, we sought to investigate *Psdr1* expression in mouse testis and liver cell lines representing distinct cell types. Steroidogenesis occurs in testicular Leydig cells, and Sertoli cells mediate the effects of these locally synthesized androgens on germ cell development. Northern blot analysis of *Psdr1* expression in the TM3 (mouse testis, Leydig cells), TM4 (mouse testis,

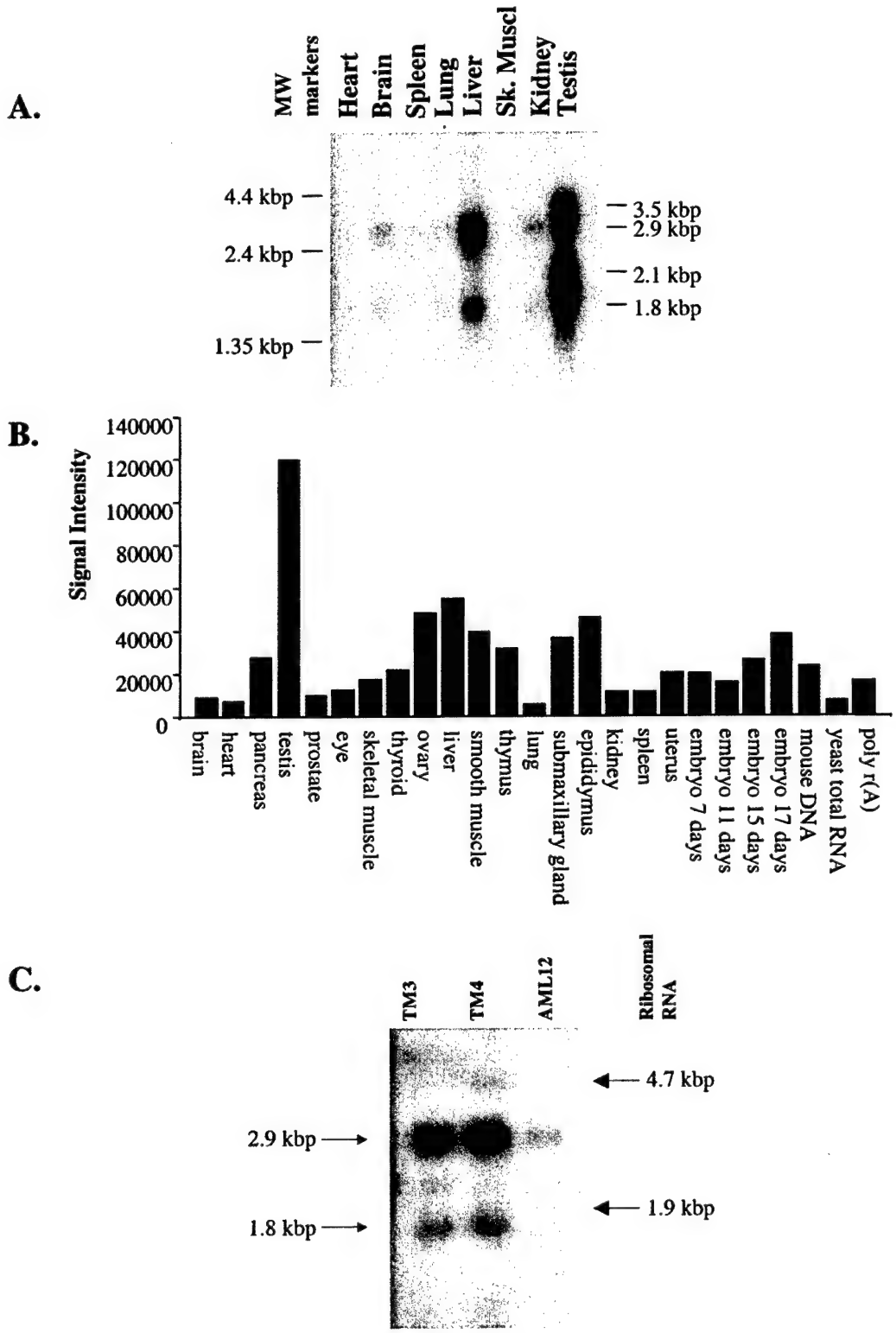


Fig. 5. Expression profile of *Psdr1* in normal mouse tissue. (A) Northern blot analysis (Clontech multiple tissue northern blot) demonstrating *Psdr1* expression and transcript size in normal mouse tissue. (B) A master tissue dot blot (Clontech) containing 22 normal mouse tissues. Signal intensities were captured with a phosphor screen, scanned with a phosphorimager, and calculated with the ImageQuant program. (C) Northern blot analyses of *Psdr1* expression in the mouse testis cell lines, TM3 (leydig) and TM4 (sertoli), and in the mouse hepatocyte cell line, AML12.

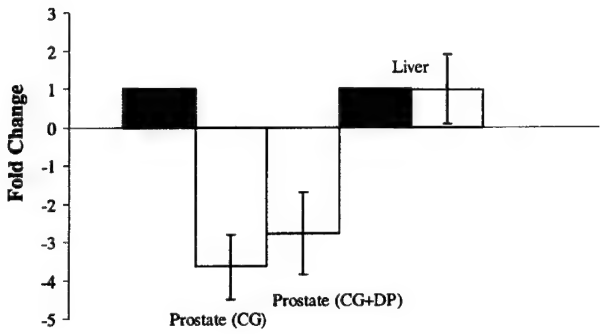


Fig. 6. Real-time PCR analysis of *Psdr1* expression in mouse prostate and liver. Expression of *Psdr1* was determined by SYBR Green real-time PCR as described in Section 2.8. The fold change of *Psdr1* expression in pooled ($n = 4$) castrate mouse prostate [coagulating gland (CG) or coagulating gland and dorsal prostate lobe (DP) combined] and liver tissue is compared to pooled control ($n = 4$) mouse prostate and liver. *Psdr1* expression was normalized to S16 expression and results are expressed as fold-change in normalized expression levels in castrate mice (open bars) relative to those in control mice (solid bars). Bars represent the average of three separate RT-PCR experiments, each performed in triplicate, with error bars representing the error of the mean of the three experiments.

Sertoli cells), and AML12 (mouse hepatocytes) cell lines demonstrated the expression of transcripts approximately 2.9 and 1.8 kbp is size (Fig. 5C). As detailed above, we hypothesize that these transcripts represent alternative splice forms of *Psdr1* or alternative use of the polyadenylation signals. Both TM3 and TM4 cells have been shown to express the androgen receptor (AR) (Mather, 1980; Chang et al., 1995; Nakhla et al., 1989a,b; Zaia et al., 2001) suggesting that expression of *Psdr1* may be regulated through AR signaling. The regulation of SDR family members involved in androgen metabolism through an AR-mediated mechanism has precedence (Couture et al., 1993). However, TM3 and TM4 cells exposed to the synthetic androgen R1881 did not upregulate *Psdr1* expression (data not shown). It is possible that the regulation of *Psdr1* expression requires a more complex signaling mechanism influenced by surrounding cell types, other endocrine factors, or nuclear receptor co-regulatory proteins not expressed in these cells. If *Psdr1* is involved in androgen metabolism, other metabolites may mediate *Psdr1* expression. Alternatively, the function of *Psdr1* may be involved in metabolic pathways distinct from androgen biosynthesis such as the modulation of retinoids. Further speculation will await the enzymatic characterization of the Psdr1 protein.

3.6. Psdr1 expression in castrated and control mouse prostate and liver

To determine whether *Psdr1* expression levels in the prostate would change in response to androgen deprivation, four C57BL/6 mice were castrated and RNA was extracted from prostate and liver for real-time PCR analysis of *Psdr1* expression. Four sham-operated C57BL/6 mice were used as controls. Real-time reverse transcription (RT)-PCR analysis demonstrated that *Psdr1* levels in the either the coagulating gland or in coagulating gland plus dorsal prostate, pooled from either castrated or control mice, was 3–4-fold lower in castrate mice relative to control mice (Fig. 6). In contrast, there was no discernable change in *Psdr1* expression in liver following castration relative to *Psdr1* expression in control liver (Fig. 6). *Psdr1* expression was normalized to the expression of a control gene, ribosomal gene S16. Expression of S16 varied by less than one cycle (i.e. less than 2-fold) when comparing prostate or liver from control versus castrate mice, indicating that changes in *Psdr1* expression are not solely a result of global decreases in gene expression which may result from castration.

These data suggest that *Psdr1* may be androgen regulated in the mouse prostate while being regulated through different mechanisms in liver tissue. Similar differential regulation of an androgen responsive gene, the AR, has been observed in the rat in which differences exist in AR mRNA regulation within the different regions of the rat prostate gland (Prins and Woodham, 1995). There are a number of levels of complexity by which differential gene regulation by androgens may occur in different tissues and in different cell types (McPhaul and Young, 2001). It is possible that *Psdr1* expression changes in response to local androgen concentration and that this response may vary depending on the tissue type.

4. Conclusions

The biological role of *Psdr1* is not known, though the homology of Psdr1 to HSD members of the SDR family suggests that PSDR1 plays a role in the metabolism of retinoids or steroid hormones (Peltoketo et al., 1999).

The putative promoter region of *Psdr1*, like that of human *PSDR1*, has predicted ARE and PRE sequences, as well as an IL6-RE BP sequence.

Comparison of the putative translations of mouse and human *PSDR1* sequences demonstrates an identity of 85%.

Phylogenetic studies with human and mouse PSDR1 protein sequences suggests that these genes are orthologous and, of the sequences analyzed, most similar to mouse 17β-HSD 7.

Chromosome mapping localizes *Psdr1* to mouse chromosome 12q31–34, a region with synteny to the human *PSDR1* chromosomal location.

Psdr1 expression, as determined by Northern blot analyses, was predominant in the testis, with transcripts also expressed at lower levels in the liver, smooth muscle, epididymus, thymus, and prostate. Analyses of mouse testis and liver cell lines indicate that *Psdr* is expressed in both the leydig and sertoli cells of the testis, as well as in hepatocytes.

Psdr1 expression decreased in mouse prostate, but not liver, in response to castration.

Acknowledgements

We thank Victor Ng for assistance with phylogenetic analyses. This work was supported in part by NIH grants CA75173 and DK59125 and Department of Defense grant DAMD PC991274 to P.S.N. and a Basic Sciences Training in Urology (DK07779-03) to S.M.M.

References

Aziz, N., Maxwell, M.M., St Jacques, B., Brenner, B.M., 1993. Downregulation of Ke 6, a novel gene encoded within the major histocompatibility complex, in murine polycystic kidney disease. Mol. Cell. Biol. 13, 1847–1853.

Bacich, D.J., Pinto, J.T., Tong, W.P., Heston, W.D., 2001. Cloning, expression, genomic localization, and enzymatic activities of the mouse homolog of prostate-specific membrane antigen/NAALADase/folate hydrolase. Mamm. Genome 12, 117–123.

Chang, C., Saltzman, A., Yeh, S., Young, W., Keller, E., Lee, H.J., Wang, C., Mizokami, A., 1995. Androgen receptor: an overview. Crit. Rev. Eukaryot. Gene Expr. 5, 97–125.

Chen, Z., Lee, W.R., Chang, S.H., 1991. Role of aspartic acid 38 in the cofactor specificity of *Drosophila* alcohol dehydrogenase. Eur. J. Biochem. 202, 263–267.

Cols, N., Marfany, G., Atrian, S., Gonzalez-Duarte, R., 1993. Effect of site-directed mutagenesis on conserved positions of *Drosophila* alcohol dehydrogenase. FEBS Lett. 319, 90–94.

Couture, P., Theriault, C., Simard, J., Labrie, F., 1993. Androgen receptor-mediated stimulation of 17 beta-hydroxysteroid dehydrogenase activity by dihydrotestosterone and medroxyprogesterone acetate in ZR-75-1 human breast cancer cells. Endocrinology 132, 179–185.

de Launoit, Y., Adamski, J., 1999. Unique multifunctional HSD17B4 gene product: 17beta-hydroxysteroid dehydrogenase 4 and D-3-hydroxyacyl-coenzyme A dehydrogenase/hydratase involved in Zellweger syndrome. J. Mol. Endocrinol. 22, 227–240.

Devgan, S.A., Henderson, B.E., Yu, M.C., Shi, C.Y., Pike, M.C., Ross, R.K., Reichardt, J.K., 1997. Genetic variation of 3 beta-hydroxysteroid dehydrogenase type II in three racial/ethnic groups: implications for prostate cancer risk. Prostate 33, 9–12.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package), 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA.

Foley, K.P., Leonard, M.W., Engel, J.D., 1993. Quantitation of RNA using the polymerase chain reaction. Trends Genet. 9, 380–385.

Geissler, W.M., Davis, D.L., Wu, L., Bradshaw, K.D., Patel, S., Mendonca, B.B., Elliston, K.O., Wilson, J.D., Russell, D.W., Andersson, S., 1994. Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3. Nat. Genet. 7, 34–39.

Hocke, G.M., Barry, D., Fey, G.H., 1992. Synergistic action of interleukin-6 and glucocorticoids is mediated by the interleukin-6 response element of the rat alpha 2 macroglobulin gene. Mol. Cell. Biol. 12, 2282–2294.

ISREC Bioinformatics Groups, Hofman, K., Bucher, P., Falquet, L., Bairoch, A., 1999. The PROSITE database, its status in 1999. Nucleic Acids Res. 27, 215–219.

Jacquinet, E., Rao, N.V., Rao, G.V., Hoidal, J.R., 2000. Cloning, genomic organization, chromosomal assignment and expression of a novel mosaic serine proteinase: epitheliasin. FEBS Lett. 468, 93–100.

Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., Ghosh, D., 1995. Short-chain dehydrogenases/reductases (SDR). Biochemistry 34, 6003–6013.

Kozak, M., 1987. An analysis of 5′-non-coding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148.

Krook, M., Marekov, L., Jornvall, H., 1990. Purification and structural characterization of placental NAD(+)-linked 15-hydroxyprostaglandin dehydrogenase. The primary structure reveals the enzyme to belong to the short-chain alcohol dehydrogenase family. Biochemistry 29, 738–743.

Lieberman, B.A., Bona, B.J., Edwards, D.P., Nordeen, S.K., 1993. The constitution of a progesterone response element. Mol. Endocrinol. 7, 515–527.

Lin, B., White, J.T., Ferguson, C., Wang, S., Vessella, R., Bumgarner, R., True, L.D., Hood, L., Nelson, P.S., 2001. Prostate short-chain dehydrogenase reductase 1 (PSDR1): a new member of the short-chain steroid dehydrogenase/reductase family highly expressed in normal and neoplastic prostate epithelium. Cancer Res. 61, 1611–1618.

Mather, J.P., 1980. Establishment and characterization of two distinct mouse testicular epithelial cell lines. Biol. Reprod. 23, 243–252.

McPhaul, M.J., Young, M., 2001. Complexities of androgen action. J. Am. Acad. Dermatol. 45, S87–S94.

Nakhla, A.M., Bardin, C.W., Salomon, Y., Mather, J.P., Janne, O.A., 1989a. The actions of calcitonin on the TM3 Leydig cell line and on rat Leydig cell-enriched cultures. J Androl. 10, 311–320.

Nakhla, A.M., Mather, J.P., Jane, O.A., Bardin, C.W., 1989b. The action of calcitonin on the TM4 Sertoli cell line and on rat Sertoli cell-enriched cultures. J. Androl. 10, 321–331.

Nokelainen, P., Puranen, T., Peltoketo, H., Orava, M., Vihko, P., Vihko, R., 1996. Molecular cloning of mouse 17 beta-hydroxysteroid dehydrogenase type 1 and characterization of enzyme activity. Eur. J. Biochem. 236, 482–490.

Norman, A.W., Litwack, G., 1997. Hormones, Academic Press, San Diego.

Oppermann, U.C., Filling, C., Berndt, K.D., Persson, B., Benach, J., Ladenstein, R., Jornvall, H., 1997a. Active site directed mutagenesis of 3 beta/17 beta-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions. Biochemistry 36, 34–40.

Oppermann, U.C., Persson, B., Jornvall, H., 1997b. Function, gene organization and protein structures of 11beta-hydroxysteroid dehydrogenase isoforms. Eur. J. Biochem. 249, 355–360.

Peltoketo, H., Luu-The, V., Simard, J., Adamski, J., 1999. 17beta-hydroxysteroid dehydrogenase (HSD)/17-ketosteroid reductase (KSR) family; nomenclature and main characteristics of the 17HSD/KSR enzymes. J. Mol. Endocrinol. 23, 1–11.

Pietrokovski, S., Henikoff, J.G., Henikoff, S., 1996. The Blocks database – a system for protein classification. Nucleic Acids Res. 24, 197–200.

Prins, G.S., Woodham, C., 1995. Autologous regulation of androgen receptor messenger ribonucleic acid in the separate lobes of the rat prostate gland. Biol. Reprod. 53, 609–619.

Pritchard, C.C., Hsu, L., Delrow, J., Nelson, P.S., 2001. Project normal: defining normal variance in mouse gene expression. Proc. Natl. Acad. Sci. USA 98, 13266–13271.

Reese, M., 1998. Neural Network Promoter Predictions. Berkeley Drosophila Genome Project, Berkeley, CA.

Roche, P.J., Hoare, S.A., Parker, M.G., 1992. A consensus DNA-binding site for the androgen receptor. Mol. Endocrinol. 6, 2229–2235.

Ross, S., Spencer, S.D., Lasky, L.A., Koeppen, H., 2001. Selective expression of murine prostate stem cell antigen in fetal and adult tissues and the transgenic adenocarcinoma of the mouse prostate model of prostate carcinogenesis. Am. J. Pathol. 158, 809–816.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schug, J.A.O., Christian, G., 1997. 'TESS: Transcription Element Search Software on the WWW', Technical Report CBIL-TR-1997-1001-v0.0. Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania, Philadelphia, PA.

Swofford, D., 1998. Phylogenetic Analysis Using Parsimony, 4.0. Sinauer Associates, Sunderland, MA.

Zaia, A., Fraizer, G.C., Piantanelli, L., Saunders, G.F., 2001. Transcriptional regulation of the androgen signaling pathway by the Wilms' tumor suppressor gene WT1. Anticancer Res. 21, 1–10.

Zhou, Z., Speiser, P.W., 1999. Regulation of HSD17B1 and SRD5A1 in lymphocytes. Mol. Genet. Metab. 68, 410–417.

The program of androgen-responsive genes in neoplastic prostate epithelium

Peter S. Nelson*†, Nigel Clegg*, Hugh Arnold*, Camari Ferguson*, Michael Bonham*, James White‡, Leroy Hood‡, and Biaoyang Lin‡

*Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024; and ‡Institute for Systems Biology, Seattle, WA 98109

Contributed by Leroy Hood, June 24, 2002

The human prostate gland is an important target organ of androgenic hormones. Testosterone and dihydrotestosterone interact with the androgen receptor to regulate vital aspects of prostate growth and function including cellular proliferation, differentiation, apoptosis, metabolism, and secretory activity. Our objective in this study was to characterize the temporal program of transcription that reflects the cellular response to androgens and to identify specific androgen-regulated genes (ARGs) or gene networks that participate in these responses. We used cDNA microarrays representing about 20,000 distinct human genes to profile androgen-responsive transcripts in the LNCaP adenocarcinoma cell line and identified 146 genes with transcript alterations more than 3-fold. Of these, 103 encode proteins with described functional roles, and 43 represent transcripts that have yet to be characterized. Temporal gene expression profiles grouped the ARGs into four distinct cohorts. Five uncharacterized ARGs demonstrated exclusive or high expression levels in the prostate relative to other tissues studied. A search of available DNA sequence upstream of 28 ARGs identified 25 with homology to the androgen response-element consensus-binding motif. These results identify previously uncharacterized and unsuspected genes whose expression levels are directly or indirectly regulated by androgens; further, they provide a comprehensive temporal view of the transcriptional program of human androgen-responsive cells.

The androgenic hormones testosterone and dihydrotestosterone exert their cellular effects by means of interactions with the androgen receptor (AR), a member of the family of intracellular steroid hormone receptors that function as ligand-dependent transcription factors (1). Ligand-activated AR, complexed with coactivator proteins and general transcription factors, binds to cis-acting androgen response elements (AREs) located in the promoter regions of specific target genes and serves to activate or to repress transcription (1, 2). During human development, circulating androgens and a functional AR mediate a wide range of reversible and irreversible effects that include the morphogenesis and differentiation of major target tissues such as the prostate, seminal vesicles, and epididimus. The prostate gland has been used extensively as a model system to study androgen effects. In part, this is because of the fact that androgens promote the development and progression of prostate diseases that account for significant morbidity in the population including benign prostatic hypertrophy and prostate adenocarcinoma (2). The recognition that normal and neoplastic prostate epithelial cells depend on circulating androgens for their continued survival and growth led to the development of effective endocrine-based therapy for prostate carcinoma (3). To date, manipulating the androgen pathway by means of surgical or chemical castration remains the primary therapeutic modality for advanced prostate cancer.

In the human prostate, the AR mediates critical processes involved in the normal development, organizational structure, and mature function of the gland. During embryogenesis, the AR is expressed in mesenchymal cells of the urogenital sinus with subsequent temporal expression in prostate epithelial cells, leading to a differentiated epithelial phenotype and the production of prostate-specific proteins (4). In the mature gland, androgens promote cell division and the proliferation of prostate epithelial cells. However, androgens also seem to modulate programmed cell death and a "proliferative shut-off" function that leads to a state of cell quiescence (5, 6). Androgens regulate several aspects of prostate cellular metabolism, including lipid biosynthesis (7), and they control the production of specialized secretory proteins with prostate-restricted expression such as prostate-specific antigen (PSA; ref. 1).

The pivotal role of androgens for the regulation of distinct and diverse physiological processes in normal and neoplastic prostate cells has led to investigations designed to identify the molecular mediators of androgen action. Elegant studies have described morphological changes and gross alterations in DNA, RNA, and protein synthesis in the prostate in response to androgen manipulation (8). Our objective in this study was to characterize the temporal program of transcription that reflects the cellular response to androgens and to identify specific androgen-regulated genes (ARGs) or gene networks that participate in these responses.

Materials and Methods

Cell Culture and General Methods. DNA manipulations including transformation, plasmid preparation, gel electrophoresis, and probe labeling were performed according to standard procedures (9). Restriction and modification enzymes (Life Technologies, Rockville, MD) were used in accordance with the manufacturer's recommendations. Prostate carcinoma cell lines LNCaP, DU145, and PC3 were cultured in phenol red-free RPMI medium 1640 supplemented with 10% (vol/vol) FCS. For androgen-regulation experiments, LNCaP cells were transferred into RPMI medium 1640 with 10% (wt/vol) charcoal-stripped FCS (CS-FCS) (Life Technologies) for 24 h followed by replacement of the media with fresh CS-FCS supplemented with 1 nM of the synthetic androgen R1881 (NEN/Life Sciences Products) or ethanol vehicle control. Cells were harvested for RNA isolation at 0-, 0.6-, 1-, 2-, 4-, 6-, 8-, 12-, 24-, and 48-h time points. Total RNA was purified from experimental and control cells by using Trizol (Life Technologies) according to the manufacturer's protocol. A reference standard RNA was prepared by combining equal quantities of total RNA isolated from LNCaP, DU145, and PC3 cell lines growing at log phase. RNA derived from one single batch of reference standard was used for every microarray hybridization. Northern analysis was performed as described (10). Multitissue Northern blots were obtained from CLONTECH.

Abbreviations: AR, androgen receptor; ARE, androgen response element; PSA, prostate-specific antigen; ARG, androgen-regulated gene; PEDB, Prostate Expression DataBase.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. BM382817).

†To whom reprint requests should be addressed at: Division of Human Biology, Mailstop D4-100, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024. E-mail: pnelson@fhcrc.org.

www.pnas.org/cgi/doi/10.1073/pnas.182376299

Microarray Experiments. A nonredundant set of ≈6,400 prostate-derived cDNA clones was identified from the Prostate Expression DataBase (PEDB), a public sequence repository of expressed sequence tag data derived from human prostate cDNA libraries (11). Microarrays were constructed as described (10). PEDB microarrays were assembled in versions composed of 3,000 or 6,388 cDNAs. A second microarray was constructed in a similar fashion by using a minimally redundant set of 17,630 human cDNA I.M.A.G.E. clones (UG Build 19V5.0, plate 1–48; Research Genetics, Huntsville, AL). Labeled cDNA probes were made from 30 μg of total RNA, as described (10). Probes were hybridized competitively to microarrays under a coverslip for 16 h at 63°C. Fluorescent array images were collected for both Cy3 and Cy5 by using a GenePix 4000A fluorescent scanner (Axon Instruments, Foster City, CA), and image intensity data were extracted and analyzed by using GENEPIX PRO 3.0 microarray analysis software. Each experiment was repeated with a switch in fluorescent labels to account for dye effects.

For each experiment, each cDNA was represented twice on each slide, and the experiments were performed in duplicate to produce four data points per cDNA clone per hybridization probe. Normalization of the Cy3 and Cy5 fluorescent signal in each experiment was determined by assuming equivalent global hybridization of test and reference probes. Data were filtered to remove values from poorly hybridized cDNAs with intensity levels less than 2 SDs above the background local to each spot. Intensity ratios for each cDNA hybridized with probes derived from the experimental time points were calculated as $\log_2$ (experimental intensity/reference intensity). Intensity ratios for each cDNA at each time point were compared with the time 0 values, and gene expression differences were considered significant if at least three of the four replicate spots for a given cDNA demonstrated an average $\log_2$ ratio of >1 or <-1 (2-fold change). Data from the four replicate cDNAs for each experiment were combined, and the average ratios were used for comparative analyses. To identify genes with similar temporal changes in expression, ratio measurements were imported into the CLUSTER software package (12). The results were visualized by using the TREEVIEW program (12).

Identification of ARE Motifs. Reference sequences of 28 characterized ARGs with a temporal profile corresponding to that of PSA were obtained from the Ref_Seq database at National Center for Biotechnology Information (13) and used to query the assembled human genome sequence at the University of California, Santa Cruz (http://genome.ucsc.edu/). Approximately 3 kb of genomic sequence upstream of each mRNA sequence was obtained and used to search for similarity to the consensus ARE motif AGAACAnnnTGTTCT (TRANSFAC; http://transfac.gbf.de/TRANSFAC/). Scoring was based on the number of nucleotides in the query sequence that matched the consensus sequence by using the web-based tool PATSEARCH VI.1 (http://transfac.gbf.de/cgi-bin/patSearch/patsearch.pl). High-scoring matches that were homopolymeric in the left or right half of the consensus sequence were excluded. The sequence with the highest score was reported for each gene; only matches with at least 9 identities of the 12 consensus nucleotides were reported (75%). If more than one putative ARE was identified, the motif mapping nearest to the 5′ end of the reference cDNA sequence was reported. All of the putative ARE motifs were aligned by using the online CLUSTALW server at the Baylor College of Medicine (http://searchlauncher.bcm.tmc.edu/). The sequence logos representing the ARE consensus sequences were generated by using the online WEBLOGO sequence generation software (www.bio.cam.ac.uk/seqlogo/). The logo characters represent the sequence as stacked nucleotide residues for each position in the aligned sequences. The height of each letter is proportional to the nucleotide frequency at each position, and the nucleotides are sorted so that the most common one is on top. The height of the entire stack is then adjusted to signify the information content of the sequences at that position. The sequence logo represents the consensus sequence, the relative frequency of bases, and the information content (measured in bits) at every position in a site or sequence (14).

Results and Discussion

Construction of a Human Prostate cDNA Microarray. cDNA libraries were produced from a variety of normal and neoplastic prostate tissue sources that included the LNCaP prostate cancer cell line. Clones were randomly selected, subjected to single-pass sequencing to generate expressed sequence tags (15), and assembled into distinct clusters based on nucleotide homology (11). Individual cDNAs corresponding to each of the ≈6,300 putative unique transcripts were selected, relocated into 384-well microtiter plates, amplified by PCR, and arrayed onto glass slides in duplicate (PEDB-Array). For some experiments described in this report, arrays consisting of a subset of 3,000 prostate cDNAs were used. For other experiments, the PEDB array was supplemented with a microarray (RG-Array) constructed with 17,630 commercially available cDNAs (Research Genetics). Overall, there are ≈3,000 genes overlapping on both arrays.

Androgen-Mediated Alterations in Gene Expression. To assess the transcriptional response of prostate epithelial cells to androgen, hormone-responsive LNCaP prostate cancer cells were exposed to the synthetic androgen R1881 for specific time periods. The LNCaP cell line was chosen because it is one of the most widely used models for the study of prostate carcinoma and of the direct effects of androgens on human cells (16). Overall, 4,439 of 6,388 genes on the PEDB-Array (69%) and 5,642 of 17,630 genes on the RG-Array (32%) exhibited detectable transcripts in the LNCaP cells for a total assessable LNCaP transcriptome representing ≈8,000 genes (the two clone sets had ≈2,000 detectable cDNAs in common). A comparison of the expression profiles at specific time points after androgen stimulation demonstrated that the vast majority of transcripts (>96%) did not change by more than 2-fold compared with untreated cells. In contrast, 3.7% of the expressed transcripts were reproducibly altered more than 2-fold at one or more time points. After 24 h of androgen stimulation, the expression of 262 genes changed by >2-fold; 183 genes increased >2-fold, and 79 genes decreased >2-fold. After either 24 or 48 h of androgen exposure, the expression of 146 genes changed by ≥3-fold; 119 transcripts increased 3-fold, and 27 transcripts decreased 3-fold. Of these, 102 are genes with described functional roles (see Table 1, which is published as supporting information on the PNAS web site, www.pnas.org), and 46 represent previously uncharacterized transcripts or putative proteins. These findings support the results from a recent report describing the use of Serial Analysis of Gene Expression to identify ARGs in prostate cells at one time point 24 h after stimulation with 10 nM R1881 (17). Of approximately 15,000 expressed genes assayed, 2.3% (351 distinct transcripts) were found to be either induced or repressed by androgen.

To determine whether androgen exposure induced distinct temporal patterns of gene expression, we used microarrays composed of 3,000 prostate-derived cDNAs to identify transcripts with a ≥2-fold change and grouped the resulting expression profiles of characterized genes by using hierarchical clustering methods (Fig. 1*A*; ref. 12). Four distinct clusters emerged from this analysis, with the largest group representing a cohort with members whose expression levels gradually increased from 4 h through 48 h (cluster C). This cohort includes the vast majority of genes previously shown to be androgen-regulated in prostate epithelium, such as KLK3/PSA, KLK2, NKX3A, TMPRSS2, TMEPA1, and SPAK (Fig. 1*B*; refs. 17–19). Other

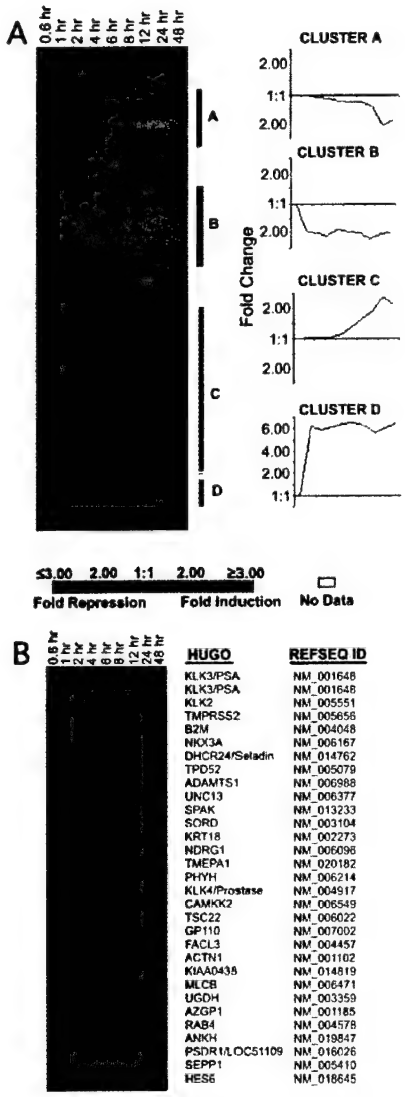


Fig. 1. Temporal expression profiles of ARGs. (*A*) Microarrays composed of 3,000 prostate-derived cDNAs were used to acquire serial measurements of androgen-induced transcripts in the LNCaP cell line. RNAs showing at least a 2-fold change in expression after androgen exposure were clustered. Groups of genes with similar patterns of expression are indicated by vertical bars (A–D). Tick marks on the *x* axis of clusters A–D temporal profiles indicate the same time intervals as depicted at *Left*. (*B*) The expanded cohort of characterized genes with a temporal profile of expression corresponding to PSA are shown and named according to the HUGO gene nomenclature (www.gene.ucl.ac.uk/nomenclature/).

distinct clusters were composed of genes with a gradual decrease in transcript levels (cluster A), genes with a rapid decrease in transcript levels (cluster B), and genes with a rapid increase in transcript levels (cluster D). Genes comprising these clusters are listed on the PEDB web site (www.pedb.org/AR/microarray).

Androgen-Regulated Expression of Characterized Genes. ARGs encoding proteins with defined biochemical function(s) were grouped into categories that reflect common functional attributes (Table 1). For clarity, each gene was assigned to only one category, although several of these genes exhibit activities that could be assigned to more than one functional role. The characteristics of the ARGs listed attest to the diverse cellular processes influenced by activation of the AR. In this discussion, we highlight several genes not previously described in the context of the cellular androgen regulation.

Androgen-Responsive Genes Involved in Metabolism. Androgens induced the expression of transcripts encoding a diverse group of proteins involved in cellular metabolism. It has previously been shown that testosterone regulates genes involved in lipid and fatty acid biosynthesis through a coordinated indirect mechanism that involves the intermediary SREBP transcription factors (20). ARGs mediating fatty acid metabolism include fatty acid synthase and acetyl-CoA-carboxylase (20). ARGs mediating cholesterol metabolism include HMG-CoA-synthase and HMG-CoA-reductase (20). Our microarray studies identified additional ARGs involved in these pathways, including stearoyl-CoA desaturase, an enzyme that functions in the synthesis of unsaturated fatty acids, HELO1, a homolog of yeast long-chain polyunsaturated fatty acid elongation enzyme 2, and long-chain fatty acid CoA ligase 3, an enzyme that converts free long-chain fatty acids into fatty acyl-CoA esters. Of further interest, transcripts encoding 3-β-hydroxysterol-Δ-24 reductase (DHCR24) were increased 4.5- and 6-fold after 24 and 48 h of androgen exposure, respectively (Fig. 2*A* and Table 1). DHCR24 is a member of the FAD-dependent oxidoreductase family and catalyzes a reduction of the Δ^{24} double bond of sterol intermediates during cholesterol biosynthesis. DHCR24 is also known as seladin1, a gene shown to be down-regulated in the affected temporal cortex of patients with Alzheimer's disease (21). Expression of the DHCR24 protein protects cells from oxidative stress and amyloid-β peptide-induced apoptosis (21).

The elevated expression of enzymes involved in lipid and cholesterol metabolism may simply reflect the mitogenic or secretory stimulus produced by androgen exposure. Cell division requires the biosynthesis of cell membranes, and the specialized secretory function of prostate epithelial cells requires the synthesis of storage vesicles and secretory components. However, there is emerging evidence that cholesterol and fatty acid metabolizing enzymes and their substrates may play more direct roles in carcinogenesis. Cholesterol seems to be intimately linked with signaling through the Ras pathway (22). Fatty acids also have been identified as signaling molecules which can be recognized by nuclear receptors (23). Numerous studies have reported an association with FAS expression and clinically aggressive cancers; one study correlates high levels of FAS expression with relapse risk in primary prostate carcinoma (24).

Androgen-Responsive Genes Involved in Transport or Trafficking. The transcript encoding the FK-506 binding-protein FKBP5 (alias FKBP51) was up-regulated 25-fold in LNCaP cells after 48 h of androgen exposure. FKBP5 is a member of the immunophilin protein family and is involved in protein folding and trafficking. Of interest in the context of hormone-mediated gene expression is a report describing a role for FKBP5 in the earliest known event in glucocorticoid-receptor signaling through participation in the control of receptor subcellular localization and transport (25). To our knowledge, specific FKBP5 interactions with the AR or AR coregulatory proteins have not been described. However, FKBP5 has been shown to be up-regulated in xenograft models of androgen-independent prostate cancer and, thus, also may participate in AR signaling (26).

Among the genes involved in processes of cellular transport, we observed a 6-fold increase in ANKH gene expression (Fig. 2*A* and Table 1). ANKH encodes a multipass transmembrane protein that regulates the transport of pyrophosphate from the cytoplasm to the extracellular space (27). Mice with mutations

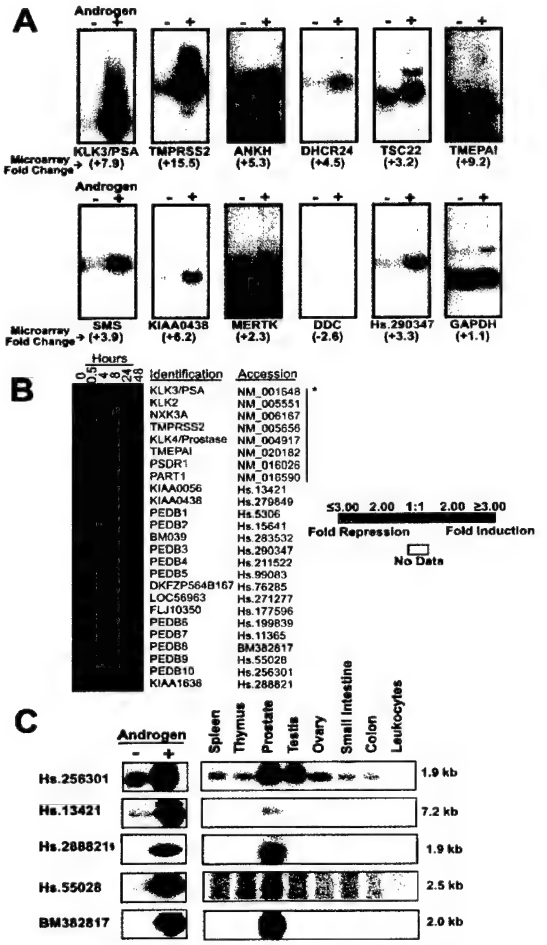


Fig. 2. Androgen-regulated expression of characterized and previously uncharacterized genes. (*A*) Northern analysis confirmation of ARG expression in LNCaP prostate cancer cells treated with the synthetic androgen R1881 (+) or vehicle control (−) for 24 h. For each gene, the corresponding microarray-derived fold change in expression is provided below the gene name. (*B*) Hierarchical cluster analysis of uncharacterized genes exhibiting a temporal expression profile corresponding to PSA. *, genes to the left of the vertical bar represent characterized ARGs with tissue-expression profiles enhanced or restricted to the prostate. (*C*) Northern analysis confirmation of uncharacterized ARGs in LNCaP cells treated with androgen (+) or vehicle control (−) for 24 h. Multiple tissue Northern blot of selected uncharacterized ARGs demonstrating prostate-restricted or prostate-enhanced expression relative to other normal human tissues. §, one alternative spliced form of Hs.288821 exhibits prostate-enhanced expression.

in the ANKH gene develop a progressive form of arthritis accompanied by calcium phosphate mineral deposition, tissue calcification, joint destruction, and the formation of bony outgrowths (27). One unique hallmark of metastatic prostate cancer is a strong tropism for bone with the development of osteoblastic lesions characterized by excessive, disorganized deposition of new bone. It is possible that ectopic or altered ANKH expression in metastatic prostate cells or the surrounding bone stromal environment could contribute to the skeletal pathology that predominate in advanced prostate carcinoma.

Androgen-Responsive Genes Involved in Cell Proliferation or Differentiation. Androgens mediate the disparate functions of prostate epithelial cell proliferation and differentiation (28). Prostate morphogenesis and overall glandular growth depends on circulating androgens during development. Subsequently, androgens maintain the differentiated functional state of the mature gland. Castration leads to the loss of differentiated secretory epithelium in rodent and human prostate tissues. This epithelium can be renewed by the restoration of androgens, but the overall proliferative response is limited to a specified cell mass through mechanisms yet to be characterized. There is evidence from several model systems that androgens also may participate in the negative regulation of cell proliferation (6, 29). A recent study using immortalized nontumorigenic rat prostate cells demonstrated a marked suppression of epithelial cell growth after cellular exposure to androgens that reflected changes in cell morphology consistent with terminal differentiation (30). The signaling mechanism(s) responsible for these distinct cellular responses have not been defined and may, in part, be temporally regulated during development and be dependent upon the cellular context of supporting stromal elements.

In this study, we identified several ARGs that encode proteins involved in cell-cycle regulation or cellular differentiation. Transcripts encoding the Maf oncoprotein were increased 16-fold after 48 h of androgen exposure. The original member of the Maf protein family (v-Maf) was identified as the transduced transforming component of avian musculoaponeurotic fibrosarcoma virus, AS42. Overexpression of Maf has been reported in multiple myeloma (31) and in melanoma cells (32). Functionally, several classes of transcriptional regulators have been shown to interact with Maf and/or Maf family proteins, including the bZip transcription factors Jun, Fos, and Bach1 (33). In addition to a role in oncogenesis, Maf mediates differentiation programs in specific cell types such as monocytes and helper T cells (34). It is hypothesized that Maf and related family members form a network with other classes of transcription factors that allow for the combinatorial fine tuning of regulatory protein interactions that dictate cellular responses that either prohibit or promote specific differentiation or growth programs.

The inhibitor of differentiation 2 (ID2) gene encodes a helix–loop–helix protein that is down-regulated in senescent human fibroblasts and during the differentiation process of lymphocyte development (35). ID2 disrupts the antiproliferative effects of the retinoblastoma protein family and negates the effect of the growth inhibitory protein p16 (36). Androgen induced the expression of ID2 3.5-fold after 48 h of exposure, an event that would be expected to relieve an inhibitory checkpoint for cellular proliferation.

Transcripts encoding several genes with reported roles in cell-cycle regulation were altered by androgen. The expression of cell division cycle 14B (Cdc14B) was up-regulated 3-fold at the 48-h time point. Cdc14 is essential for cell-cycle progression in yeast and encodes a protein tyrosine phosphatase involved in the exit of cell mitosis and initiation of DNA replication. In contrast to ID2 and Cdc14B, androgen decreased the expression of cyclin-dependent kinase 8 (CDK8). CDK8 and its regulatory subunit cyclin C are components of the RNA polymerase II holoenzyme complex which phosphorylates the largest subunit of RNA polymerase II (RNAPII). The cell cycle and transcription by RNAPII are closely related, and yeast orthologs of CDK8 and cyclin C have been implicated in the negative regulation of transcription. The mechanism of CDK8/cyclin C transcriptional regulation involves phosphorylation of the CDK7/cyclin H subunits of the general transcription initiation factor IIH (TFIIH), which results in repression of both kinase activity and the ability to activate transcription (37). Mimicking CDK8 phosphorylation of cyclin H *in vivo* has a dominant–negative effect on cell growth. Combined, the effect of androgen to down-regulate negative regulators of the cell cycle and induce positive regulators serves to promote a proliferative response. Future work should determine whether these molecular alter-

ations represent specific androgenic cellular responses or rather reflect a more general mitogenic stimulus.

Androgen-Regulated Expression of Uncharacterized Genes. In addition to genes with defined cellular functions, we identified 46 ARGs with homology only to hypothetical proteins or uncharacterized expressed sequence tags. Our criteria for prioritizing ARGs for verification and biochemical characterization centered on those with a tissue-distribution profile enhanced or specific to prostate tissue. Such genes may provide biological insights into unique facets of prostate physiology, or they may provide diagnostic or therapeutic targets for the treatment of prostate carcinoma. As a first step toward identifying genes with enhanced expression in the prostate, we selected ARGs with a temporal pattern of expression similar to that of PSA, a gene shown to be directly regulated by androgens and used extensively in clinical applications as a diagnostic and prognostic marker of prostate carcinoma. Seventeen uncharacterized ARGs grouped with KLK3/PSA and other well described prostate ARGs, such as KLK2 and NKX3A (Fig. 2*B*). We further investigated the tissue distribution of these ARGs by Northern analysis by using RNAs representing eight different human tissues (Fig. 2*C*). Five genes exhibited exclusive or high expression levels in the prostate relative to all other tissues studied, and these were confirmed to be induced by androgens in the LNCaP cell line (Fig. 2*C*). KIAA0056, originally identified through large-scale sequencing of cDNA clones from the immature myeloid cell line KG-1, encodes a putative protein of 1,498 residues (38). The Unigene sequence represented by Hs.256301 encodes the putative protein MGC13170 cloned from a retinoblastoma cDNA library. Interestingly, this sequence maps to chromosome 19q13.3, a region harboring several androgen-regulated prostate proteases including PSA/KLK3, KLK2, and KLK4/prostase. The protein predicted to be encoded by Hs.288821 does not exhibit overall similarity with any characterized human protein but does have WD-domains and has significant homology to proteins predicted in the *Drosophila melanogaster* and *Caenorhabditis elegans* genomes. By Northern analysis, several isoforms of Hs.288821 were shown to be present in human tissues with one isoform showing prostate specificity (Fig. 2*C*). The cDNA corresponding to Unigene sequence Hs.55028 maps to chromosome 16 and encodes a predicted polypeptide of 33 amino acids with similarity to a protein candidate for X-linked retinopathies (GenBank accession no. A46010). The cDNA for PEDB8 was derived from a library constructed from normal prostate tissue and does not exhibit significant homology with sequences in the public nucleotide databases.

Regulation of Androgen-Responsive Gene Expression. Androgenic hormones exert their biological effects through the regulated expression of specific effector proteins and the initiation of signaling cascades. One mechanism of androgen-mediated gene expression involves the direct interaction of hormone with the AR protein, resulting in nuclear translocation and interactions with specific DNA sequences located near or within androgen target genes (1). Binding to these promoter and enhancer sequences, known as AREs and androgen regulatory regions, facilitates interactions with the general transcriptional machinery leading to gene transcription. Androgens also can affect gene expression through posttranscriptional (39) and genome-independent mechanisms (40). Alternatively, androgen target genes may be regulated indirectly: as a secondary or tertiary event through the initial direct up-regulation or liberation of a transcription factor(s) that in turn regulates the expression of other target genes (20). Such a network allows for layers of regulatory control that may be advantageous for the temporal direction of protein synthesis, the amplification of androgen

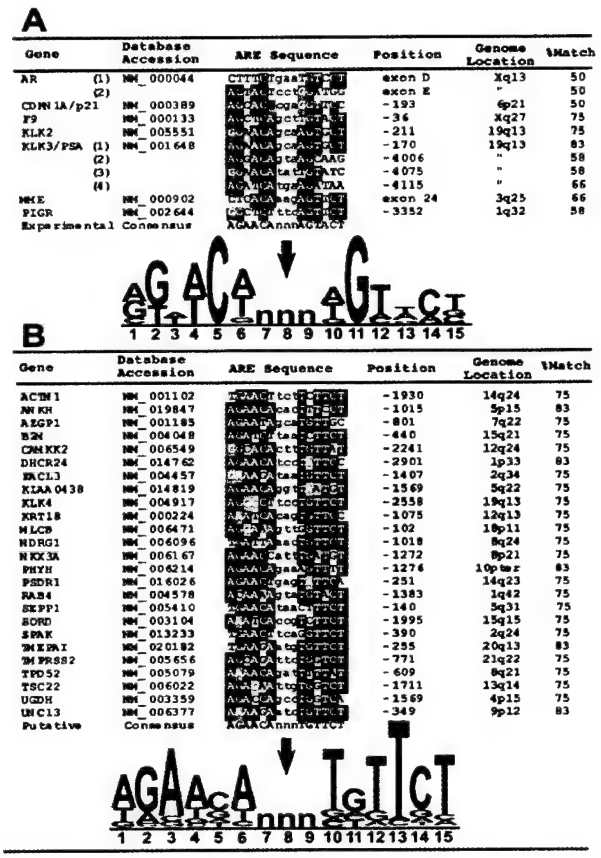


Fig. 3. Identification of ARE motifs. (*A*) Functional human AREs verified through experimentation are shown with positions relative to the transcriptional start site and the approximate genome location. A CLUSTALW alignment identifies highly conserved residues in black. A consensus sequence generated by WEBLOGO displays the frequency of each base in the consensus proportional to the character height with the height of the entire stack adjusted to signify the information content of the sequences at that position. (*B*) Putative human AREs identified by searching the 5′ regulatory regions of androgen target genes for a motif corresponding to the Transfac ARE consensus sequence. A CLUSTALW alignment identifies highly conserved residues in black. A consensus sequence indicates the relative frequency and importance of nucleotides in the motif.

signaling, and the coordinated expression of genes involved in common metabolic processes.

To gain an understanding of potential regulatory mechanisms operative in the ARGs identified in this study, we sought to identify sequences with similarity to known AREs that could support a mechanism of direct, rather than indirect, transcriptional control. Most AREs conform to a consensus sequence composed of two 6-base asymmetrical elements separated by three spacer nucleotides; 5′-AGAACAnnnTGTTCT-3′ (http://transfac.gbf.de/TRANSFAC/). To date, AREs in seven human genes have been characterized experimentally by others using reporter gene and gel-shift experiments (Fig. 3*A*). The AR and PSA genes have been shown to contain multiple functional AREs (41, 42). Importantly, operational human AREs that do not conform to the consensus mammalian ARE sequence have been described (42). A CLUSTALW alignment of the experimentally confirmed human AREs diverges from the consensus ARE with particular variability in positions 3 and 13 of the 15-nucleotide motif (Fig. 3*A*).

To obtain DNA sequences containing putative gene regulatory elements, we searched the assembled human genome with mRNA reference sequences encoded by the 28 characterized ARGs with temporal expression profiles corresponding to that of PSA (Fig. 1*B*). Approximately 3 kb of sequence upstream of the putative transcriptional start sites were examined for homology to the consensus ARE obtained from the TRANSFAC database of eukaryotic cis-acting regulatory DNA elements. We identified 25 genes containing a motif comprising at least 9 of 12 nucleotides corresponding to the consensus ARE (Fig. 3*B*). A CLUSTALW alignment of these putative AREs suggests a high conservation of the "right-half" sequence, TGTTCT. Direct repeats of this motif have been shown to confer high-affinity AR binding that may contribute to androgen-selective responses *in vivo* (43). The biochemical characterization of these putative AREs should be useful for the delineation of critical nucleotides mediating DNA–AR interactions and allow for further studies of the contextual arrangement of AREs and other regulatory motifs in relation to transcriptional control.

We thank Jeff Delrow, Cassie Neal, Ryan Bosum, and Jan Kim for assistance with microarray experiments, and Mike Eisen for the public availability of microarray analysis software. This work was supported by the CaPCURE Foundation, National Cancer Institute Grants CA75173 and CA85286, and Department of Defense Grants DAMD17-98-1-8499 and DAMD17-00-1-0050 (to P.S.N.). P.S.N. is supported by a scholar award from the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation.

1. Prins, G. S. (2000) *Mayo Clin. Proc.* **75,** S32–S35.
2. Eder, I. E., Culig, Z., Putz, T., Nessler-Menardi, C., Bartsch, G. & Klocker, H. (2001) *Eur. Urol.* **40,** 241–251.
3. Huggins, C. & Hodges, C. V. (1941) *Cancer Res.* **1,** 293–297.
4. Bonkhoff, H. & Remberger, K. (1996) *Prostate* **28,** 98–106.
5. Isaacs, J. T., Lundmo, P. I., Berges, R., Martikainen, P., Kyprianou, N. & English, H. F. (1992) *J. Androl.* **13,** 457–464.
6. Geck, P., Szelei, J., Jimenez, J., Lin, T. M., Sonnenschein, C. & Soto, A. M. (1997) *J. Steroid Biochem. Mol. Biol.* **63,** 211–218.
7. Swinnen, J. V. & Verhoeven, G. (1998) *J. Steroid Biochem. Mol. Biol.* **65,** 191–198.
8. Bruchovsky, N., Lesser, B., Van Doorn, E. & Craven, S. (1975) *Vitam. Horm. (San Francisco)* **33,** 61–102.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
10. Lin, B., Ferguson, C., White, J. T., Wang, S., Vessella, R., True, L. D., Hood, L. & Nelson, P. S. (1999) *Cancer Res.* **59,** 4180–4184.
11. Hawkins, V., Doll, D., Bumgarner, R., Smith, T., Abajian, C., Hood, L. & Nelson, P. S. (1999) *Nucleic Acids Res.* **27,** 204–208.
12. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14863–14868.
13. Pruitt, K. D. & Maglott, D. R. (2001) *Nucleic Acids Res.* **29,** 137–140.
14. Schneider, T. D. & Stephens, R. M. (1990) *Nucleic Acids Res.* **18,** 6097–6100.
15. Nelson, P. S., Ng, W. L., Schummer, M., True, L. D., Liu, A. Y., Bumgarner, R. E., Ferguson, C., Dimak, A. & Hood, L. (1998) *Genomics* **47,** 12–25.
16. Horoszewicz, J. S., Leong, S. S., Kawinski, E., Karr, J. P., Rosenthal, H., Chu, T. M., Mirand, E. A. & Murphy, G. P. (1983) *Cancer Res.* **43,** 1809–1818.
17. Xu, L. L., Su, Y. P., Labiche, R., Segawa, T., Shanmugam, N., McLeod, D. G., Moul, J. W. & Srivastava, S. (2001) *Int. J. Cancer* **92,** 322–328.
18. Nelson, P. S., Han, D., Rochon, Y., Corthals, G. L., Lin, B., Monson, A., Nguyen, V., Franza, B. R., Plymate, S. R., Aebersold, R. & Hood, L. (2000) *Electrophoresis* **21,** 1823–1831.
19. Qi, H., Labrie, Y., Grenier, J., Fournier, A., Fillion, C. & Labrie, C. (2001) *Mol. Cell. Endocrinol.* **182,** 181–192.
20. Swinnen, J. V., Ulrix, W., Heyns, W. & Verhoeven, G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12975–12980.
21. Greeve, I., Hermans-Borgmeyer, I., Brellinger, C., Kasper, D., Gomez-Isla, T., Behl, C., Levkau, B. & Nitsch, R. M. (2000) *J. Neurosci.* **20,** 7345–7352.
22. Cox, A. D. & Der, C. J. (1992) *Crit. Rev. Oncog.* **3,** 365–400.
23. Vanden Heuvel, J. P. (1999) *J. Nutr.* **129,** 575S–580S.
24. Shurbaji, M. S., Kalbfleisch, J. H. & Thurmond, T. S. (1996) *Hum. Pathol.* **27,** 917–921.
25. Davies, T. H., Ning, Y. M. & Sanchez, E. R. (2001) *J. Biol. Chem.* **20,** 20.
26. Amler, L. C., Agus, D. B., LeDuc, C., Sapinoso, M. L., Fox, W. D., Kern, S., Lee, D., Wang, V., Leysens, M., Higgins, B., *et al.* (2000) *Cancer Res.* **60,** 6134–6141.
27. Ho, A. M., Johnson, M. D. & Kingsley, D. M. (2000) *Science* **289,** 265–270.
28. Davies, P. & Eaton, C. L. (1991) *J. Endocrinol.* **131,** 5–17.
29. Sato, N., Gleave, M. E., Bruchovsky, N., Rennie, P. S., Goldenberg, S. L., Lange, P. H. & Sullivan, L. D. (1996) *J. Steroid Biochem. Mol. Biol.* **58,** 139–146.
30. Whitacre, D. C., Chauhan, S., Davis, T., Gordon, D., Cress, A. E. & Miesfeld, R. L. (2002) *Cell Growth Differ.* **13,** 1–11.
31. Chesi, M., Bergsagel, P. L., Shonukan, O. O., Martelli, M. L., Brents, L. A., Chen, T., Schrock, E., Ried, T. & Kuehl, W. M. (1998) *Blood* **91,** 4457–4463.
32. Li, M., Huang, X., Zhu, Z. & Gorelik, E. (1999) *J. Virol.* **73,** 9178–9186.
33. Kataoka, K., Shioda, S., Yoshitomo-Nakagawa, K., Handa, H. & Nishizawa, M. (2001) *J. Biol. Chem.* **276,** 36849–36856.
34. Ho, I. C., Hodge, M. R., Rooney, J. W. & Glimcher, L. H. (1996) *Cell* **85,** 973–983.
35. Yokota, Y., Mansouri, A., Mori, S., Sugawara, S., Adachi, S., Nishikawa, S. & Gruss, P. (1999) *Nature (London)* **397,** 702–706.
36. Lasorella, A., Iavarone, A. & Israel, M. A. (1996) *Mol. Cell. Biol.* **16,** 2570–2578.
37. Akoulitchev, S., Chuikov, S. & Reinberg, D. (2000) *Nature (London)* **407,** 102–106.
38. Nomura, N., Nagase, T., Miyajima, N., Sazuka, T., Tanaka, A., Sato, S., Seki, N., Kawarabayasi, Y., Ishikawa, K. & Tabata, S. (1994) *DNA Res.* **1,** 251–262.
39. Perry, J. E. & Tindall, D. J. (1996) *Cancer Res.* **56,** 1539–1544.
40. Kousteni, S., Bellido, T., Plotkin, L. I., O'Brien, C. A., Bodenner, D. L., Han, L., Han, K., DiGregorio, G. B., Katzenellenbogen, J. A., Katzenellenbogen, B. S., *et al.* (2001) *Cell* **104,** 719–730.
41. Cleutjens, K. B., van Eekelen, C. C., van der Korput, H. A., Brinkmann, A. O. & Trapman, J. (1996) *J. Biol. Chem.* **271,** 6379–6388.
42. Dai, J. L. & Burnstein, K. L. (1996) *Mol. Endocrinol.* **10,** 1582–1594.
43. Claessens, F., Verrijdt, G., Schoenmakers, E., Haelens, A., Peeters, B., Verhoeven, G. & Rombauts, W. (2001) *J. Steroid Biochem. Mol. Biol.* **76,** 23–30.

Appendix
DAMD17-00-1-0050
Peter S. Nelson, PI

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ACADEMIC PRESS

GENOMICS

Genomics 0 (2003) 000–000

www.elsevier.com/locate/ygeno

Isolation and characterization of human and mouse WDR19, a novel WD-repeat protein exhibiting androgen-regulated expression in prostate epithelium☆

Biaoyang Lin,[a,*] James T. White,[a] Angelita G. Utleg,[a] Shunyou Wang,[b] Camari Ferguson,[c] Lawrence D. True,[d] Robert Vessella,[b] Leroy Hood,[a] and Peter S. Nelson[c]

[a] *The Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103, USA*
[b] *Department of Urology, University of Washington, Seattle, WA 98195, USA*
[c] *Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA*
[d] *Department of Pathology, University of Washington, Seattle, WA 98195, USA*

Received 30 April 2002; accepted 13 May 2003

Abstract

Androgens regulate important processes involved in the normal development and function of the human and rodent prostate glands. Here we report the isolation and characterization of a new androgen-regulated gene, designated *WDR19,* that encodes repeating sequence motifs found in the WD-repeat family of proteins. The WD repeat is a conserved domain of approximately 40 amino acids that is typically bracketed by glycine–histidine and tryptophan–aspartic acid (WD) dipeptides. WD-repeat proteins are a large group of structurally related proteins that participate in a wide range of cellular functions, including transmembrane signaling, mRNA modification, vesicle formation, and vesicular trafficking. The *WDR19* gene comprises 36 exons and is located on chromosome 4p15–4p11. The predicted protein contains six WD repeats, a clathrin heavy-chain repeat, and three transmembrane domains. Sequence analysis reveals that the *WDR19* gene is conserved from *Caenorhabditis elegans* to human. *WDR19* is expressed in normal and neoplastic prostate epithelium as demonstrated by RNA in situ hybridization and is regulated by androgenic hormones. *WDR19* transcripts exhibit alternative splicing in which two isoforms appear to be prostate restricted, a property that could be exploited for designing diagnostic or therapeutic strategies for prostate carcinoma.

Keywords: WD repeat; WDR; Androgen; Clathrin; Cloning; Prostate; Alternative splicing

Androgenic hormones are important mediators of normal urogenital development and serve to maintain the male phenotype. Androgens are also implicated in the development and progression of prostate adenocarcinoma. We have developed strategies directed toward the identification of androgen-regulated genes in the prostate as a first step toward understanding the role of androgenic hormones in normal prostate function and in pathological conditions affecting the prostate gland. In this study we characterized the prostate cellular transcriptional response to androgens using a custom-made microarray comprising ~6000 cDNAs derived specifically from prostate cDNA libraries (www.pedb.org) [1,2]. One of the androgen-regulated transcripts encodes a previously undescribed protein predicted to contain six tryptophan–aspartic acid (WD)[1] dipeptide repeats. We have designated this human gene and corresponding mouse ortholog *WDR19* and *Wdr19,* respectively.

The WD-repeat protein (WDR) family comprises a large group of functionally distinct but structurally related proteins that contain a minimally conserved repetitive sequence

☆ Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AY029257 (human) and AY029258 (mouse).

* Corresponding author. Fax: +1-206-732-1299.
E-mail address: blin@systemsbiology.org (B. Lin).

[1] Abbreviations used: ARE, androgen-responsive element; CHCR, clathrin heavy-chain repeat; CS, charcoal-stripped; EST, expressed sequence tag; PSA, prostate-specific antigen; WDR, WD-repeat-containing protein; RACE, rapid amplification of cDNA ends; WD, tryptophan–aspartic acid.

0888-7543/03/$ – see front matter
doi:10.1016/S0888-7543(03)00151-4

of approximately 40 amino acids that is typically bracketed by a glycine–histidine dipeptide at the N-terminus and a tryptophan–aspartic acid (WD) dipeptide at the C-terminus [3]. WD-repeat proteins usually contain at least four WD repeats [4]. These multiple WD domains can form a donut-like structure or β propeller that produces a scaffold for protein–protein interactions [3]. To date, 123 WD-repeat-containing proteins have been characterized with entries in the Swiss-Prot/TrEMBL databases (http://bmerc-www.bu.edu/wdrepeat/sw-34_n_sp.html).

WD-repeat-containing proteins participate in a broad spectrum of cellular functions such as gene transcription [5], apoptosis [6], cytoskeletal assembly, mitotic-spindle formation [7], development [8], and vesicular trafficking (SEC13) [9]. For example, several transmembrane signal transduction proteins belong to the WD-repeat protein family, including the Gβ subunit of retinal G protein transducin [10]; RACK1, an anchoring protein for activated PKC [11]; and TRIP-1, a protein associated with the type II TGF-β receptor [12]. Recently, mutations in WD-repeat proteins have been implicated in diseases such as X-linked sensorineural deafness (OMIM 300650), characterized by ocular albinism and progressive and late onset sensorineural hearing loss [13]; the Cockayne syndrome (OMIM 216400), characterized by dwarfism and mental retardation [14]; and the triple A syndrome (OMIM 231550), characterized by achalasia, alacrima, and adrenocorticotropin hormone insensitivity [15,16].

The *WDR19* gene reported here comprises 36 exons spanning a 110-kb genomic region that maps to chromosome 4p15–4p11. The predicted protein contains six WD repeats, a clathrin heavy-chain repeat (CHCR), and three transmembrane domains. Sequence analysis reveals that the *WDR19* gene is conserved from *Caenorhabditis elegans* to humans. *WDR19* is expressed in normal and neoplastic prostate epithelium as demonstrated by RNA in situ hybridization. Alternative splicing of the *WDR19* transcript produces two isoforms that appear to be prostate restricted. Prostate epithelial cells carry out specialized functional roles involving the production and secretion of seminal fluid proteins such as prostate-specific antigen (PSA), an activity that is under the control of androgens. Thus, based on the attributes of other WD-repeat-containing proteins, the normal role for WDR19 may be to participate in androgen-regulated signaling mechanisms or in the vesicular trafficking of androgen-regulated secretory processes.

Results

Identification, cloning, and sequence analysis of a novel human androgen-regulated gene: WDR19

We used cDNA microarray analysis to profile androgen-induced gene expression alterations in the androgen-sensitive prostate adenocarcinoma cell line LNCaP [17]. A previously uncharacterized cDNA clone, later designated WDR19 after

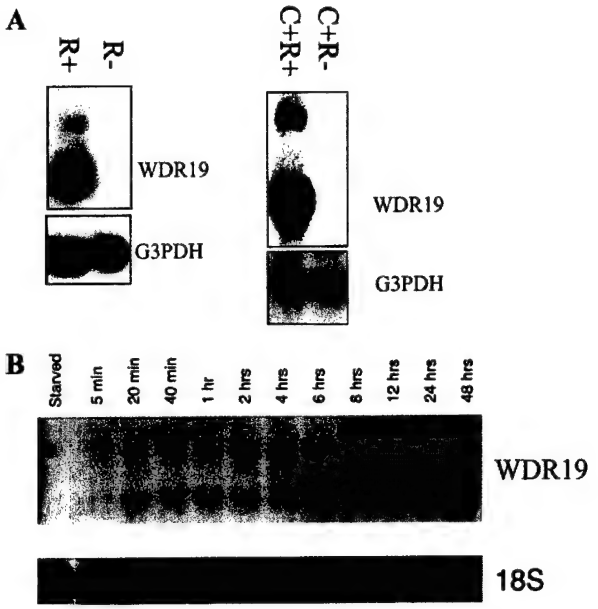


Fig. 1. Northern analysis of WDR19 expression. (A) Left: WDR19 expression in androgen-stimulated (+) (24 h) and androgen-starved (−) LNCaP cells. Right: WDR19 expression in androgen-stimulated (+) (24 h) and androgen-starved (−) LNCaP cells in the presence of cycloheximide. (B) Time-course analysis of WDR19 expression in LNCaP cells 5, 20, and 40 min and 1, 2, 4, 8, 12, 16, 24, and 48 h after androgen stimulation.

cloning and sequence analysis, increased threefold in androgen-stimulated LNCaP cells relative to androgen-deprived cells. Northern analysis confirmed that *WDR19* expression is induced by androgen following 24 h of androgen exposure (Fig. 1A, left). The induction of *WDR19* expression by androgens was not inhibited by cycloheximide (Fig. 1A, right). Furthermore, the regulation of *WDR19* expression was rapid, as an increase in *WDR19* message levels was detected as early as 20 min following androgen exposure and subsequently increased gradually with time (Fig. 1B).

To clone the full-length WDR19 cDNA, a human prostate 5′-STRETCH cDNA library was used to screen for additional WDR19 cDNA clones using the original WDR19 partial-length clone as a probe. The 5′ end of WDR was cloned by 5′ RACE. The full-length WDR19 cDNA is 4410 nucleotides and encodes a predicted protein of 1342 amino acid residues (Fig. 2). The sequence was deposited with GenBank under Accession No. AY029257. An ATG start codon (GCC*ATG*G) that conforms to the Kozak translation initiation consensus sequence (RNN*ATG*G, where R is a purine) [18] was identified at nucleotide position 332. This start codon also aligns with the only strong ATG start codon identified in the murine *Wdr19* ortholog (described below), suggesting that this codon is probably the major human WDR19 start codon. A weak ATG (GAG*ATG*A) start codon was also identified upstream at nucleotide position 155. Although a weak ATG codon can also be an authentic site of initiation of translation [18], further studies are needed to

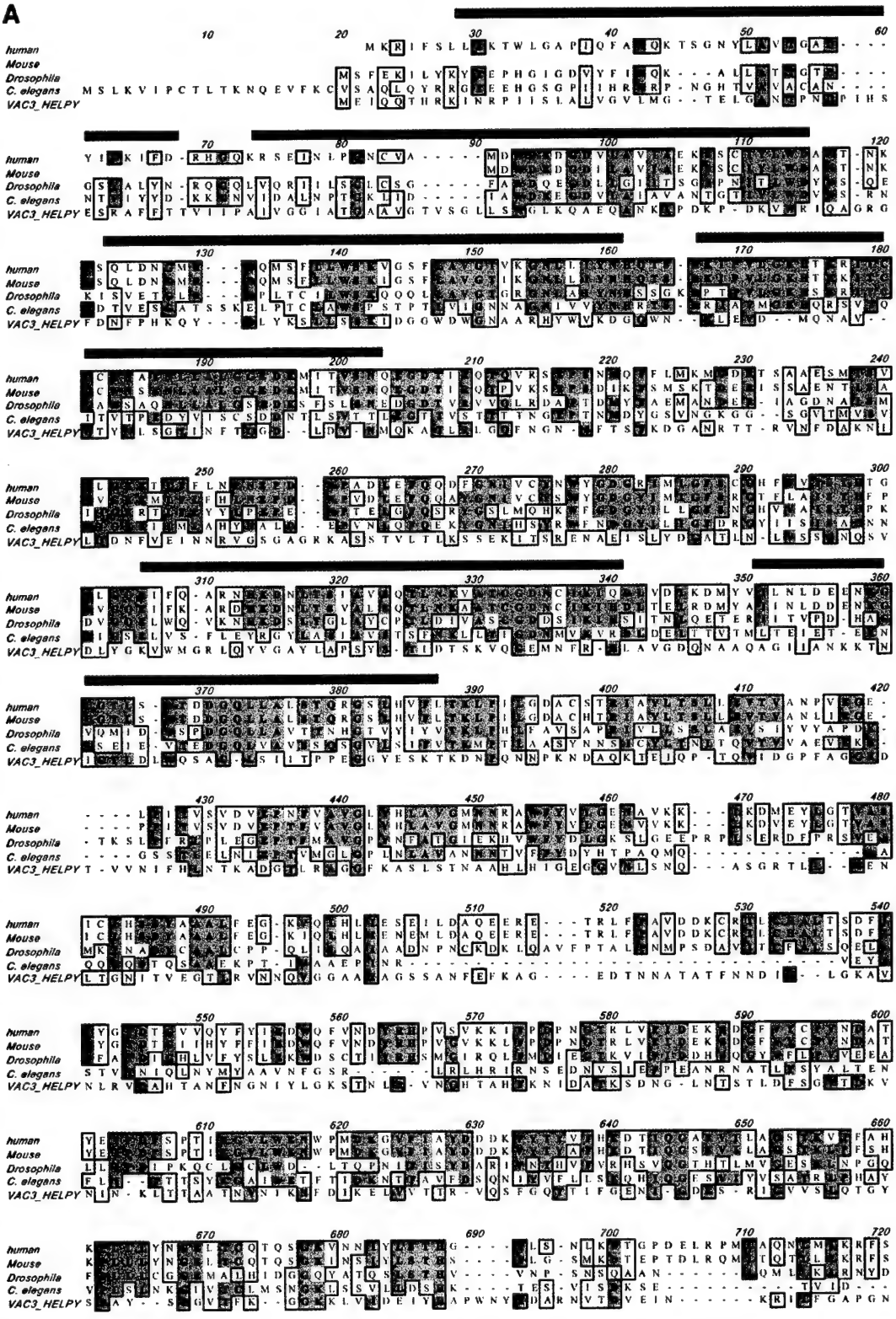


Fig. 2. WDR19 protein alignments. ClustalW alignment of WDR19 orthologs from human (GenBank: AY029257), mouse (GenBank: AY029258), *Drosophila* (GenBank: AAF57545), *C. elegans* (T27881), and a vacuolating cytotoxin protein (VACA-HELPY, Swiss-Prot Q48253) from *Helicobacter pylori*. Identical amino acid residues are boxed and dark-shaded. Similar amino acid residues are boxed and light-shaded. A thick line on the top of the amino acid residues indicates the WD repeat.

Fig. 2(*continued*)

determine whether this start codon is used. A polyadenylation site, AATAAA, was identified at nucleotide position 4393, 18 bp 5′ of the poly(A) tail.

The predicted WDR19 protein sequence was examined for conserved regions that could indicate functional attributes. Sequence comparisons with the pfam database of protein motifs (http://pfam.wustl.edu/) [19] indicate that WDR19 contains six WD repeats (Fig. 2). This result was confirmed using the Biomolecular Engineering Research Center search engine using a WD-repeat consensus sequence (http://bmerc-www.bu.edu/wdrepeat/sw-34_n_sp. html) [3]. Beyond the WD domain, WDR19 does not exhibit homology with other recently characterized human WD-repeat-containing proteins, WDR1–13 and WDR16–18 [20–23], or with 123 WD-repeat-containing proteins identified in the Swiss-Prot/TrEMBL database (http://bmerc-www.bu.edu/wdrepeat/sw-34_n_sp.html). The predicted WDR19 protein also has homology to a CHCR over amino acid residues 818–970. Prosite searches identified 4 potential N-glycosylation sites, 3 cAMP- and cGMP-dependent protein kinase phosphorylation sites, 16 protein kinase C phosphorylation sites, 15 casein kinase II phosphorylation sites, 4 tyrosine kinase phosphorylation sites, 17 N-myristoylation sites, and 1 amidation site in the WDR19 protein. The significance of these motifs remains to be determined. Transmembrane domain analysis with the Tmpred program [24] predicts that the WDR19 protein has three strong transmembrane helices derived from amino acid residues 102–124, 243–259, and 390–412, indicating that it is a type III membrane protein.

Cloning of the mouse Wdr19

Searches against the mouse Expressed Sequence Tag (EST) database with the human *WDR19* sequence identified three distinct clusters of ESTs with significant homology to human *WDR19*. The first cluster with homology to the 3′ end of the human WDR19 cDNA is represented by mouse ESTs AI505067, AW554824, and AI558190; the second cluster, with homology to the middle region of WDR19, is represented by mouse ESTs BB608262 and BF429719; and the third, close to the 5′ end of WDR19, is represented by mouse EST BB569003. Following alignment with the human sequence, two gaps of about 3.0 and 0.5 kb existed and sequence corresponding to the 5′ end was absent. PCR primers were designed and used to clone the sequence gaps, and the 5′ end of murine *Wdr19* was cloned by RACE.

The full-length mouse Wdr19 cDNA is 4471 nucleotides in length and encodes a predicted protein of 1282 amino acids. The sequence was deposited with GenBank under the Accession No. AY029258. A start codon, ACC*ATG*G, which conforms to the Kozak translation initiation consensus sequence (RNN*ATG*G, where R is a purine) [18], was identified at nucleotide position 307. Sequence comparisons with human *WDR19* demonstrate that mouse *Wdr19* exhibits 82% nucleotide identity (3742 nt of the aligned length of 4559 nt) and 89% amino acid identity (1146 of 1283 amino acids). A polyadenylation site, AATAAA, is located at nucleotide position 4446, 26 nt upstream of the poly(A) tail.

WDR19 is an evolutionarily conserved protein

Searches against the protein sequence databases revealed that WDR19 exhibits homology with GenBank protein entries zk520.1 (CAB07299) and zk520.3 (CAB07301) from *C. elegans* and protein entry AAF57545 from *Drosophila melanogaster*. Sequence alignments produced using ClustalW (MacVector, Oxford Molecular Group) indicate that *C. elegans* CAB07301 aligns with the 5′ half of the human WDR19 protein and CAB07299 aligns with the 3′ half of the WDR19 protein, suggesting that CAB07301 and CAB07299 actually originate from one gene. Indeed, CAB07299 and CAB07301 are predicted proteins from one single genomic region, zk520, on *C. elegans* chromosome III. WDR19 displays 30% amino acid identity (425 of the aligned length of 1398) and 48% amino acid similarity with the combined CAB07299 and CAB07301 protein sequences. The human WDR19 protein has 40% amino acid identity (555 of the aligned length of 1419 amino acids) and 62% amino acid residue similarity with the putative *D. melanogaster* protein. A ClustalW alignment of the human WDR19 protein with its murine and putative *Drosophila* and *C. elegans* orthologs is shown in Fig. 2.

The human WDR19 and its murine ortholog *Wdr19* protein are predicted to contain six WD repeats. Similarly, six WD repeats were found in the putative *Drosophila* WDR19 ortholog and four WD repeats were identified in the *C. elegans* WDR19 ortholog. The CHCR domain contained in human WDR19 is also conserved in these WDR19 orthologs (data not shown). The homology between these sequences extends beyond the WD- and CHCR-repeat regions, suggesting that they are true orthologs of each other. Unfortunately, no functional information about the encoded WDR19 *C. elegans* and *Drosophila* proteins has been reported.

A sequence of 1172 amino acids (aa 2–1174) of the WDR19 protein exhibits homology with the vacuolating cytotoxin protein VAC3_HELPY from *Helicobacter pylori* [25]. This cytotoxin can induce cytoplasmic vacuolation in a number of different mammalian cell lines [25]. Alignment of the VAC3_HELPY protein with the *C. elegans, Drosophila,* mouse, and human WDR19 protein sequences reveals numerous conserved blocks and 38 amino acids that are identical in respective position in the five aligned proteins (Fig. 2), suggesting that the vacuolating cytotoxin protein may be a distant relative of WDR19. A search for protein sorting signals with pSORT II (http://psort.nibb.ac.jp/) revealed that the WDR19 protein has a vacuolar targeting motif (KLPI) at amino acid residue 389.

WDR19 chromosomal localization and genomic organization

The medium-resolution Stanford G3 radiation hybrid panel was used to map the chromosomal location of

WDR19. Analysis of the PCR results on the SHGC RH server (www.shgc.stanford.edu) indicated that *WDR19* is localized to SHGC-8532 between two cytogenetically mapped markers, D4S756 (mapped to 4p14–4p11) and D4S174 (mapped to 4p15–4p11) (http://www.gdb.org/). Therefore, *WDR19* maps to chromosome 4p14–4p11. Two bacterial artificial chromosomes (BAC) that contain the *WDR19* gene (GenBank Accession Nos. AC018858 and AC023135) mapped to the 34–39 Mb interval on chromosome 4, consistent with our mapping result.

Blat searches against the assembled human genome sequence (genome.ucsc.edu) revealed that *WDR19* spans a 110-kb genomic region and contains 36 exons in which all intron/exon junctions conform to the GT–AG rule (data not shown). A genomic region located 4 kb upstream of the first *WDR19* exon exhibits significant homology to ESTs AI377320, AI299803, and AI301782, which encode the 3′ end of a different gene, suggesting a boundary for the 5′ *WDR19* gene terminus. Blat searches with the murine *Wdr19* sequence against the mouse genome revealed that *Wdr19* also contains 36 exons in which all intron/exon junctions conform to the GT–AG rule (data not shown). The *Wdr19* genomic region is at least 60 kb. A more definitive size determination awaits the joining of gaps in several *Wdr19* introns.

A search for promoter elements located within 4 kb of the putative *WDR19* translation initiation codon was performed using the Signal Scan software tool (http://bimas.dcrt.nih.gov/molbio/signal/). An AP1 site at position −361, an SP1 site at −166, and a TATA box at position −754 relative to the *WDR19* ATG start codon were identified. Three sequences exhibiting at least 75% homology with the consensus androgen-response element (ARE), 5′-GGA/TACAnnnTGTTCT-3′ [26], were identified. These are located at the nucleotide positions −643 (ACAACAaaaTGTGCT), −1526 (AGAGCAatcAGTTCT), and −1738 (GGTTCAcgcCGTTCT) relative to the predicted start codon.

Systematic searches of the entire 110-kb genomic region of *WDR19* were performed to identify additional ARE sequences and CAAT or TATA motifs that could represent alternative promoters. In intron 14, a putative ARE sequence (GTAACAactTGTTCT) was identified at −1982 nt 5′ of exon 15 (first nucleotide in exon 15 is +1). An AP1 site at position −485, an SP1 site at position −241, a CAAT sequence at position −294, and a TATA sequence at position −31 were identified within this intron, suggesting that intron 14 could be used as alternative promoter. A transcript derived from the use of intron 14 promoter elements would generate a 2.9-kb transcript, a size similar to the 3.0-kb form seen by Northern analysis. In intron 27, a putative ARE sequence (CATCCAtccTGTTCT) was identified −694 nt 5′ of exon 28 and a TATA sequence was located at −512 nt 5′ of exon 28, suggesting that this intron could also be an alternative promoter and would generate a transcript of about 1.3 kb (Fig. 3). Proteins encoded by the use of these alternative promoters would not contain the six WD repeats of WDR19 and would not include the transmembrane domains.

WDR19 also exhibits alternate splicing and differential usage of polyadenylation sites. For example, a search of the EST database identified an EST (AW450839) that contains an additional exon of 313 nucleotides (exon 21A) after exon 21 (Fig. 3). This additional exon ends with a poly(A) tail, suggesting that the cDNA represented by EST AW450839 is derived from both alternate splicing and an alternative use of polyadenylation signals. The sequence of this alternately spliced exon aligns with the sequence from the same BAC clone, AC018858, which contains other exons of WDR19. However, a probe composed solely of exon 21A sequence failed to detect any hybridization signal on two human multiple-tissue Northern blots that included the same 16 tissues as those used in Fig. 4 (data not shown). This suggests that the transcript containing exon 21A is expressed either at low levels or in tissues not represented in our analysis. A cDNA represented by EST AW386761 will have a larger exon 33 because of alternative use of the 3′ acceptor site at intron 32 resulting in a size increase of 233 bp (exon 33A). EST BE928712 retains the sequence of intron 34 and thus joins exon 34, intron 34, and exon 35 to become one exon (exon 35A). This increases the *WDR19* transcript size by 330 nucleotides (Fig. 3).

Two alternatively spliced WDR19 transcripts are androgen-regulated and abundantly expressed in the prostate

The tissue distribution of *WDR19* transcripts was assessed by Northern and dot-blot analysis using RNAs derived from multiple human and mouse tissues. In a dot-blot analysis of 76 human tissues, the original WDR19 cDNA clone (nucleotides 3292–4517 of the full-length WDR19 cDNA) was most highly expressed in the human prostate (Fig. 4A, location E8), but was also detectable in the cerebellum, pituitary gland, fetal lung, and pancreas (Fig. 4A, locations A2, B9, D3, and G11, respectively). Northern analysis with the same sequence identified four distinct transcripts of 1.8, 3.0, 4.5, and 6.8 kb in the normal prostate (Fig. 4B). Transcripts of 4.5 and 6.8 kb were observed in the testis and ovary. A faint band corresponding to the 1.8-kb transcript was detected in the pancreas. In the androgen-responsive LNCaP prostate cancer cell line, the 4.5- and 6.8-kb transcripts were not detectable, the 1.8-kb form was highly expressed, and the 3.0-kb transcript was expressed at a low level (Fig. 4C). The expression of the 1.8- and 3.0-kb transcripts was markedly induced by androgen exposure (Fig. 4C). No detectable level of WDR19 expression was observed in prostate cancer cell lines, DU145 and PC-3, that do not express a functional androgen receptor (Fig. 4C).

To study the tissue distribution of the mouse *Wdr19* transcript, Northern blots comprising 15 normal mouse tissues were probed with a cDNA sequence corresponding to nucleotides 3688–4372 of the mouse *Wdr19* cDNA. Three bands corresponding to 7.4, 4.3, and 1.0 kb were identified in multiple tissues (Fig. 5A). The 1.0-kb transcript is the

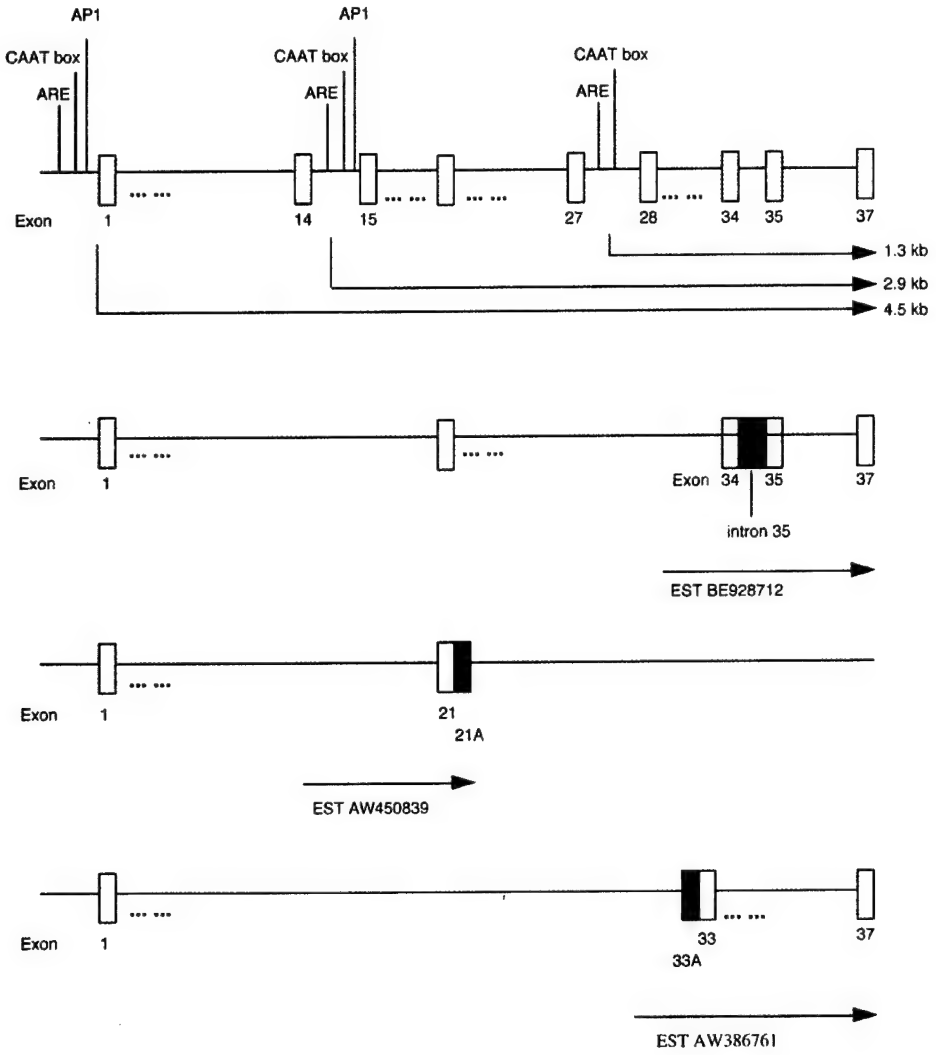


Fig. 3. Putative alternative promoter and alternative splicing sites in the *WDR19* gene.

most abundant form in the prostate and salivary glands (Fig. 5A). In testis, the most abundant form is a 1.2-kb form. Hybridization of the same probe to a mouse mRNA dot blot revealed that the additive expression levels of the three *Wdr19* transcript forms are highest in the submaxillary gland, testis, and epididymis (Fig. 5B, locations C4, D1, and D4 respectively).

WDR19 expression in normal and neoplastic prostate epithelium

The normal prostate gland comprises multiple cell types including basal epithelium, luminal secretory epithelium, smooth muscle, fibroblasts, neuroendocrine cells, and vascular endothelium. We performed in situ hybridizations on tissue sections of normal prostate and prostate adenocarcinoma to localize the cellular distribution of *WDR19* expression. Adenocarcinoma cells were uniformly positive for *WDR19* expression (Fig. 6A), and hybridizations with sense *WDR19* RNA probes showed no background staining (Fig. 6B). *WDR19* expression was detected in both normal basal and normal luminal epithelial cell populations (Fig. 6C). Little to no staining was seen in fibromuscular stromal cells, endothelial cells, or infiltrating lymphocytes. Hybridization with sense *WDR19* RNA probes showed no background staining (Fig. 6D).

Discussion

WDR19 represents a new member of the expanding family of WD-repeat-containing proteins that now encompasses more than 120 distinct constituents. However, the functions of most of these genes and their cognate proteins

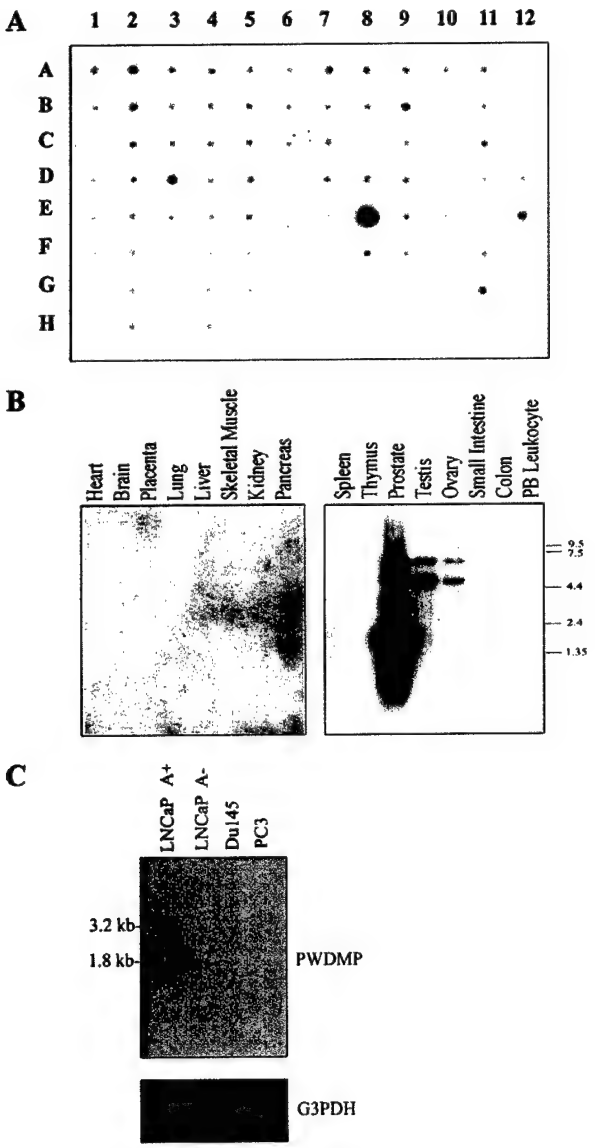


Fig. 4. Tissue distribution profile of human WDR19 transcripts. (A) Dot blot comprising RNA from 76 different human tissues hybridized with WDR19 cDNA probe (nt 3292–4410). A map of the blot tissue locations is available online (http://www.clontech.com/archive/JAN99UPD/humanmte.shtml). Spots with intense hybridization signals are human prostate (E8), left cerebellum (A2), pituitary gland (D3), fetal lung (G11), and pancreas (B9). (B) Multiple-tissue Northern analysis with a WDR19 cDNA probe (nt 3292–4410). (C) Northern analysis with RNA derived from prostate cancer cell lines PC3, Du145, and LNCaP ±24 h of androgen stimulation using a WDR19 cDNA probe (nt 3292–4410).

remain largely unknown. In addition to the WD domains, *WDR19* also encodes transmembrane-spanning regions and a CHCR. The clathrin proteins are the major proteins of the polyhedral coat of coated pits and vesicles. These proteins are involved in the transport of vesicles from the rough endoplasmic reticulum to the Golgi network and then to the plasma membrane [27,28]. The expression of clathrin heavy-chain and light-chain proteins has been shown to be regulated by androgens in the prostate gland and they are thought to be involved in the androgen-regulated secretion of PSA and other prostatic proteins [29]. The CHCR motif is also found in nonclathrin proteins such as Pep3, Pep5, Vam6, Vps41, and Vps8 from *Saccharomyces cerevisiae* and their orthologs from other eukaryotes [30]. These proteins, like clathrins, are involved in vacuolar maintenance and protein sorting [31]. Interestingly, WDR19 also contains a vacuolar-targeting motif. Thus, a possible function of the WDR19 protein could involve vacuole generation and transport.

Expression of the *WDR19* gene produces several transcriptional isoforms that exhibit tissue-selective profiles.

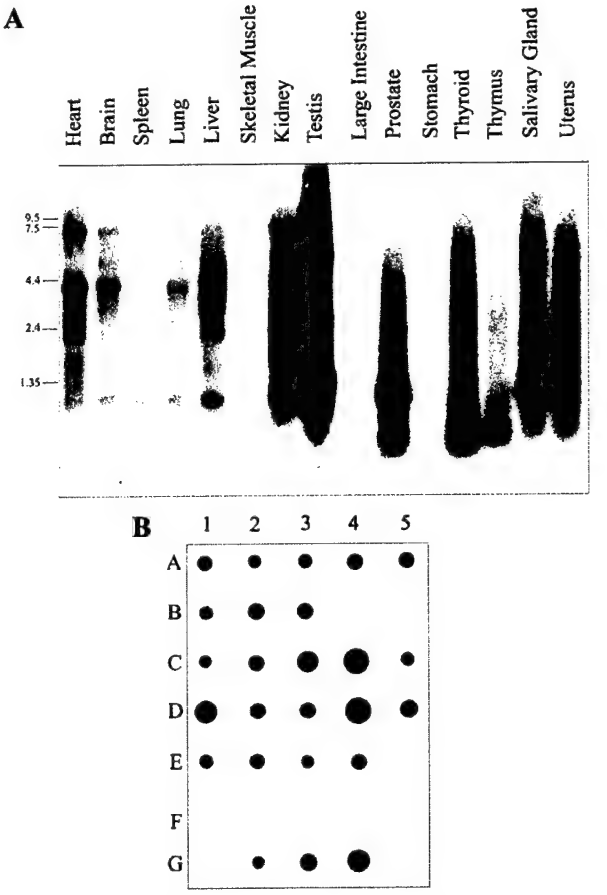


Fig. 5. Tissue distribution of mouse *Wdr19* expression. (A) Multiple-tissue Northern analysis using a mouse Wdr19 cDNA probe (nt 3292–4410). (B) Multiple-tissue RNA dot-blot analysis of mouse Wdr19 cDNA probe (nt 3292–4410). The tissue sources and blot locations for 22 different mouse RNAs are A1, brain; A2, eye; A3, liver; A4, lung; A5, kidney; B1, heart; B2, skeletal muscle; B3, smooth muscle; B4, blank; B5; blank; C1, pancreas; C2, thyroid; C3, thymus; C4, submaxillary gland; C6, spleen; D1, testis; D2, ovary; D3, prostate; D4, epididymis; D5, uterus; E1, embryo, 7 days; E2, embryo, 11 days; E3, embryo, 15 days; F1, yeast total RNA; F2, yeast tRNA; F3, *Escherichia coli* rRNA; F4, *E. coli* DNA, F5, blank; G1, poly r(A); G2, C_0t I DNA; G3 and G4, mouse DNA; G5, blank.

Fig. 6. Localization of *WDR19* expression to normal and neoplastic prostate epithelium. Representative sections of in situ hybridization with *WDR19* probes in normal and malignant prostate tissues. (A) Antisense and (B) sense *WDR19* probe hybridization to prostate adenocarcinoma showing expression in neoplastic prostate epithelium with minimal expression in stromal cells. (C) Antisense and (D) sense *WDR19* probe hybridization to normal prostate tissue showing *WDR19* expression in normal basal and secretory epithelium and minimal expression in stromal cells. Sense probes demonstrated no cross-reactivity. All sections are counterstained with hematoxylin.

Two transcripts are highly expressed in the prostate gland and these isoforms of WDR19 represent the first WDR proteins shown to be regulated by androgenic hormones. Cycloheximide experiments indicate that the induction of these WDR19 isoforms by androgen does not require de novo protein synthesis, suggesting that the expression of these transcripts is regulated directly by androgen, rather than through intermediary transcription factors. We have identified putative androgen-responsive elements and CAAT motifs in three potential alternative promoter regions of WDR19. However, final proof that these alternative promoters are used and are responsive directly to activation through the androgen receptor will require additional experimentation.

Alternative splicing and alternative use of polyadenylation sites and promoters appear to be common mechanisms regulating mammalian gene expression. Croft et al. [32] estimated that about 22% of human genes may be alternatively spliced. EST-based analyses of 475 disease-causing genes [33] reveals that one in three genes exhibits alternative splicing. Recently, the finding that the human genome has only 30,000–40,000 genes [34,35] supports a conclusion that human genes are more complex than previously

thought, with more alternative splicing to generate a larger number of proteins from a limited number of genes. These mechanisms may also explain the different *WDR19* transcripts in normal prostate and other human tissues. The 2.9- and 1.8-kb transcripts were seen in prostate tissue but not in testis and ovary. They may derive from tissue-specific alternative use of promoters. Tissue-specific alternative promoter use and alternative splicing offer a mechanism for gene regulation that can also serve to expand the potential functions of a limited repertoire of genes encoded in the genome. The glucokinase gene serves as a good example as two different promoter regions and different sets of transcription factors are used to regulate tissue-specific expression of glucokinase mRNAs in the liver, pancreatic β cell, and pituitary. This allows glucokinase gene expression to be regulated by insulin and cAMP in liver, and by glucose in the β cell, resulting in maintenance of blood glucose homeostasis [36].

Proteins expressed specifically in the prostate may provide insights into the normal specialized functions of the gland and could also be exploited for developing diagnostic and therapeutic modalities for prostate diseases. For example, the prostate-specific forms of WDR19 could potentially be targeted by immunotherapy strategies for the treatment of prostate carcinoma. Elucidation of the mechanism(s) directing the prostate-specific expression of the WDR19 isoforms could be useful for directing gene therapy approaches as has been widely pursued using PSA or prostate-specific membrane antigen (PSMA) promoter/enhancers [37–39]. Interestingly, the tissue expression patterns of *WDR19* in human and mouse are different. Genes expressed in mammalian accessory organs often show species-restricted expression patterns. For example, Aumuller et al. [40]showed that the expression profiles of semenogelin, acid phosphatase, β-microseminoprotein, and PSA is species- and organ-specific. The human PSMA (FOH1) is expressed specifically in the human prostate [41]; however, its murine homologue (*foh1*) is not expressed in the mouse prostate, but primarily in the hippocampal region of the brain and kidney [42]. Thus, extrapolating functional studies of WDR19 in model systems such as the mouse will need to be interpreted with care.

Materials and methods

Cell culture

LNCaP cells were routinely cultured in RPMI 1640 medium with 5% FBS (Life Technologies, Rockville, MD, USA) and transferred into RPMI 1640 medium with 10% charcoal-stripped FCS (CS-FCS) (Life Technologies) 48 h before androgen-regulation experiments. This medium was replaced with fresh CS-FCS medium or CS-FCS supplemented with 1 nM synthetic androgen R1881 (Perkin–Elmer, Wellesley, MA, USA). Cells were harvested for RNA isolation at 4-, 8-, 12-, 16-, 24-, 26-, and 48-h time points. 9E1_MS_3688F (CCATCACACATCGTGCCTATCC) and 9E1_MS_4372R (GAGCACAGAACACACAGGACTTTG) were used to amplify the mouse *Wdr19* 3′ portion from mouse testis marathon cDNAs (Clontech, Inc., Palo Alto, CA, USA) and used as probe for experiments shown in Fig. 5C. The PCR conditions were 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min.

Microarray fabrication, hybridization, and data analysis

Approximately 6400 cDNAs derived from the Prostate Expression Database [1] were used to construct cDNA microarrays. The microarray fabrication and hybridization protocols were performed as described previously [17].

Northern hybridization

Ten micrograms of total RNA was fractionated on 1.2% agarose denaturing gels and transferred to nylon membranes by a capillary method [43]. Human and mouse multiple-tissue and master blots were purchased from Clontech, Inc. Blots were hybridized with DNA probes labeled with [α-^{32}P]dCTP by random priming using the Rediprime II random primer labeling system (Amersham, Piscataway, NJ, USA) according to the manufacturer's protocol. Filters were imaged and quantified using a phosphor-capture screen and Imagequant software (Molecular Dynamics, Sunnyvale, CA, USA).

Library screening and rapid amplification of cDNA ends

Approximately 1.2 million phage plaques from a human prostate 5′-STRETCH cDNA library (Clontech) were screened with the inserts of WDR19 clones using standard methods [43]. After tertiary screening, the clone inserts were amplified by the PCR and directly sequenced. Clones extending the original WDR19 sequence were used again to screen the library in an iterative fashion.

Human prostate Marathon-Ready cDNA (Clontech) was used for RACE. Template cDNA was also made from androgen-stimulated LNCaP cells using the Marathon cDNA amplification kit (Clontech) according to the manufacturer's protocol. RACE primers were WDR19_RC82 (5′-CAAGGTAGTTTCCTGATGTTTTTTGCCAGG-3′) and WDR19_RC215 (5′-TCAGCAATCACTGCTAGGACATCTCCATC-3′). The RACE products were subcloned into PCR2.1-TOPO vectors with the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) and sequenced.

Cloning of mouse Wdr19

Mouse primers WDR19_MS_gap1 (5′-GAACCCTTCACCTTGGCTCAG-3′) and WDR19_MS_gap2 (5′-TGGCGATGATGATGGCGG-3′) were used to amplify the region between the first and the second EST cluster from

BALB/c mouse testis Marathon-Ready cDNA (Clontech). The PCR conditions were 35 cycles with each cycle at 94°C for 30 s, 55°C for 30 s, and 72°C for 3 min. Forward primer WDR19_MS_FL1 (5′-TGAAGCGTGTTTTCTCCCTG-3′) and forward nested primer WDR19_MS_FL2 (5′-ATTCT-TGGCTTGGTGCTCC-3′) were designed from BB569003 sequence. Reverse primer WDR19_MS_FL_R1 (5′-TCT-TCAGCACCAATGATGTCTG-3′) and nested reverse primer WDR19_MS_Fl_R2 (5′-CCATTTTGTTGTGCT-GCTGAG-3′) were designed from the middle EST cluster the mouse *Wdr19* cDNA. The PCR conditions for *Wdr19*_MS_FL1 and *Wdr19*_MS_FLR1 were 35 cycles with each cycle at 94°C for 30 s, 55°C for 30 s, and 72°C for 4 min. A band of the expected 3-kb size was amplified using the first primer pair and subjected to nested PCR using the nested primer pair. To generate a minilibrary for sequencing, the product resulting from the nested PCR was subcloned into pDNA2.1 and subjected to EZ::TN transposon insertion with EZ::TN Insertion Kits (Epicenter Technologies, Madison, WI, USA), using the manufacturer's protocol. *Pfu* DNA polymerase (Promega, Madison, WI, USA) was used for RT-PCR. The 5′ end of mouse *Wdr19* was cloned by RACE using primer *Wdr19* _MS_5RACE_1 (5′-CCAGGCTAATTGTATTGGAGCACCAAGC-3′) and nested primer *Wdr19* _MS_5RACE_2 (5′-GCACCAAGC-CAAGAATTTTATAGCAGGGAG-3′) in a PCR with BALB/c mouse testis Marathon-Ready cDNA (Clontech) according to the manufacturer's protocols.

In situ hybridization

A PCR product was generated from the 3′ end of the WDR19 sequence using primers WDR19insitu1 (5′-TGAA-GAACTCTGCTTTCAGCTTCGC-3′) and WDR19insitu2 (5′-AGGAAACAGCCTCCTGTGGAAAATG-3′). The reaction product was cloned into PCRII-TOPO (Invitrogen), linearized at either end with BamHI or EcoRV, and transcribed to generate sense and antisense digoxigenin-labeled probes according to the manufacturer's instructions (Boehringer Mannheim, Germany). In situ hybridization was performed on an automated instrument (Ventana Gen II; Ventana Medical Systems, Tucson, AZ, USA) as described in Lin et al. [44]. The hybridized probes were detected using a cocktail of anti-rabbit and anti-mouse secondary IgG biotinylated antibody with an indirect biotin avidin diaminobenzidine detection system. The sections were counterstained with hematoxylin.

Chromosomal localization of WDR19

The medium-resolution Stanford G3 radiation hybrid panel (Research Genetics, Huntsville, AL, USA) was used to map the chromosomal localization of *WDR19* with primers WDR19MapF (5′-ACGTGCAGATACAATGCTCCT-GAG-3′) and WDR19MapR (5′-CATGTCATCGTTTTGC-CACCG-3′). After 35 cycles of amplification, the reaction products were separated on a 1.2% agarose gel, and the resulting product pattern was analyzed through the Stanford Genome Center Web server (www.shgc.stanford.edu) to determine the probable chromosomal location.

Other general methods and sequence analysis

DNA manipulations including transformation, plasmid preparation, gel electrophoresis, and probe labeling were performed according to standard procedures [43]. Restriction and modification enzymes (Life Technologies) were used in accordance with the manufacturer's recommendations.

Sequence assemblies were performed using Sequencher 4.1 (Gene Codes Corp., Ann Arbor, MI, USA). General sequence analyses including conceptual translations of nucleotide sequences and ClustalW multiple sequence alignments were done using MacVector 6.5 (Accelrys, Inc., San Diego, CA, USA). Other sequence analyses such as motif searches were carried out as indicated under Results.

Acknowledgments

This work was supported by grants from the National Cancer Institute (CA75173-01A1) and Department of Defense (PC991274) to P.S.N. and a grant from the CaPCURE Foundation to L.H.

References

[1] P.S. Nelson, et al., The prostate expression database (PEDB): status and enhancements in 2000, Nucleic Acids Res. 28 (2000) 212–3.
[2] V. Hawkins, et al., PEDB: the Prostate Expression Database, Nucleic Acids Res. 27 (1999) 204–8.
[3] T.F. Smith, C. Gaitatzes, K. Saxena, E.J. Neer, The WD repeat: a common architecture for diverse functions, Trends Biochem. Sci. 24 (1999) 181–5.
[4] R. Apweiler, et al., InterPro—an integrated documentation resource for protein families, domains and functional sites, Bioinformatics 16 (2000) 1145–50.
[5] C.A. Keleher, M.J. Redd, J. Schultz, M. Carlson, A.D. Johnson, Ssn6-Tup1 is a general repressor of transcription in yeast, Cell 68 (1992) 709–19.
[6] F. Cecconi, G. Alvarez-Bolado, B.I. Meyer, K.A. Roth, P. Gruss, Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development, Cell 94 (1998) 727–37.
[7] N. Vaisman, et al., The role of Saccharomyces cerevisiae Cdc40p in DNA replication and mitotic spindle formation and/or maintenance, Mol. Gen. Genet. 247 (1995) 123–36.
[8] S. Stifani, C.M. Blaumueller, N.J. Redhead, R.E. Hill, S. Artavanis-Tsakonas, Human homologs of a Drosophila Enhancer of split gene product define a novel family of nuclear proteins [published erratum appears in Nat. Genet. Dec;2(4):(1992) 343], Nat. Genet. 2 (1992) 119–27.
[9] K. Saxena, et al., Analysis of the physical properties and molecular modeling of Sec13: a WD repeat protein involved in vesicular traffic, Biochemistry 35 (1996) 15215–21.
[10] H.K. Fong, et al., Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs, Proc. Natl. Acad. Sci. USA 83 (1986) 2162–6.

[11] M.M. Rodriguez, D. Ron, K. Touhara, C.H. Chen, D. Mochly-Rosen, RACK1, a protein kinase C anchoring protein, coordinates the binding of activated protein kinase C and select pleckstrin homology domains in vitro, Biochemistry 38 (1999) 13787–94.
[12] R.H. Chen, P.J. Miettinen, E.M. Maruoka, L. Choy, R. Derynck, A WD-domain protein that is associated with and phosphorylated by the type II TGF-beta receptor, Nature 377 (1995) 548–52.
[13] M.T. Bassi, et al., X-linked late-onset sensorineural deafness caused by a deletion involving OA1 and a novel gene containing WD-40 repeats, Am. J. Hum. Genet. 64 (1999) 1604–16.
[14] K.A. Henning, et al., The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH, Cell 82 (1995) 555–64.
[15] A. Tullio-Pelet, et al., Mutant WD-repeat protein in triple-A syndrome, Nat. Genet. 26 (2000) 332–5.
[16] K. Handschug, et al., Triple A syndrome is caused by mutations in AAAS, a new WD-repeat protein gene, Hum. Mol. Genet. 10 (2001) 283–90.
[17] B. Lin, et al., Prostate short-chain dehydrogenase reductase 1 (PSDR1): a new member of the short-chain steroid dehydrogenase/reductase family highly expressed in normal and neoplastic prostate epithelium, Cancer Res. 61 (2001) 1611–8.
[18] M. Kozak, Interpreting cDNA sequences: some insights from studies on translation, Mamm. Genome 7 (1996) 563–74.
[19] A. Bateman, et al., The Pfam protein families database, Nucleic Acids Res. 28 (2000) 263–6.
[20] H.J. Adler, R.S. Winnicki, T.W. Gong, M.I. Lomax, A gene upregulated in the acoustically damaged chick basilar papilla encodes a novel WD40 repeat protein, Genomics 56 (1999) 59–69.
[21] J.O. Claudio, et al., Cloning and expression analysis of a novel WD repeat gene, WDR3, mapping to 1p12–p13, Genomics 59 (1999) 85–9.
[22] D. Li, et al., Molecular cloning, expression analysis, and chromosome mapping of WDR6, a novel human WD-repeat gene, Biochem. Biophys. Res. Commun. 274 (2000) 117–23.
[23] Gross C., De Baere E., Lo A., Chang W., Messiaen L. (2001). Cloning and characterization of human WDR10, a novel gene located at 3q21 encoding a WD-repeat protein that is highly expressed in pituitary and testis. DNA Cell Biol. 20: 41-52.
[24] Hofmann K. and S., W. (1993). TMbase—a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 347: 166.
[25] J.C. Atherton, et al., Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori, J. Biol. Chem. 270 (1995) 17771–17777.
[26] P.J. Roche, S.A. Hoare, M.G. Parker, A consensus DNA-binding site for the androgen receptor, Mol. Endocrinol. 6 (1992) 2229–35.
[27] G. Griffiths, S. Pfeiffer, K. Simons, K. Matlin, Exit of newly synthesized membrane proteins from the trans cisterna of the Golgi complex to the plasma membrane, J. Cell Biol. 101 (1985) 949–64.
[28] L. Orci, M. Ravazzola, M. Amherdt, D. Louvard, A. Perrelet, Clathrin-immunoreactive sites in the Golgi apparatus are concentrated at the trans pole in polypeptide hormone-secreting cells, Proc. Natl. Acad. Sci. USA 82 (1985) 5385–9.
[29] J.L. Prescott, D.J. Tindall, Clathrin gene expression is androgen regulated in the prostate, Endocrinology 139 (1998) 2111–9.
[30] J.A. Ybe, et al., Clathrin self-assembly is mediated by a tandemly repeated superhelix, Nature 399 (1999) 371–5.
[31] E. Conibear, T.H. Stevens, Vacuolar biogenesis in yeast: sorting out the sorting proteins, Cell 83 (1995) 513–6.
[32] L. Croft, et al., ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome, Nat. Genet. 24 (2000) 340–1.
[33] J. Hanke, et al., Alternative splicing of human genes: more the rule than the exception? Trends Genet. 15 (1999) 389–90.
[34] J.C. Venter, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.
[35] I.H.G.S. Consortium, Initial sequencing and analysis of the human genome, Nature 409 (2001) 860.
[36] M.A. Magnuson, Tissue-specific regulation of glucokinase gene expression, J. Cell. Biochem. 48 (1992) 115–21.
[37] D. Yu, et al., Prostate-specific targeting using PSA promoter-based lentiviral vectors, Cancer Gene Ther. 8 (2001) 628–35.
[38] J.P. Latham, P.F. Searle, V. Mautner, N.D. James, Prostate-specific antigen promoter/enhancer driven gene therapy for prostate cancer: construction and testing of a tissue-specific adenovirus vector, Cancer Res. 60 (2000) 334–41.
[39] D.S. O'Keefe, et al., Prostate-specific suicide gene therapy using the prostate-specific membrane antigen promoter and enhancer, Prostate 45 (2000) 149–57.
[40] G. Aumuller, et al., Species- and organ-specificity of secretory proteins derived from human prostate and seminal vesicles, Prostate 17 (1990) 31–40.
[41] R.S. Israeli, C.T. Powell, J.G. Corr, W.R. Fair, W.D. Heston, Expression of the prostate-specific membrane antigen, Cancer Res. 54 (1994) 1807–11.
[42] D.J. Bacich, J.T. Pinto, W.P. Tong, W.D. Heston, Cloning, expression, genomic localization, and enzymatic activities of the mouse homolog of prostate-specific membrane antigen/NAALADase/folate hydrolase, Mamm. Genome 12 (2001) 117–23.
[43] J. Sambrook, E.F. Fritsch, T. Maniatis, "Molecular Cloning", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[44] B. Lin, et al., Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2, Cancer Res. 59 (1999) 4180–4.